Mental Health in Context: Effect Heterogeneity and Mechanism in the Relationship between Neighborhood Disadvantage and Adolescent Depression and Anxiety

by
Kara E. Rudolph

A dissertation submitted to The Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy.

Baltimore, Maryland
January, 2014

Abstract

Research into the association between neighborhood disadvantage and adolescent stress and anxiety/depressive disorders has produced inconsistent results. Contributing factors may include lack of generalizability and incomplete control for confounding and positivity violations due to neighborhood selection and segregation. The goal of this dissertation is to apply causal inference methods to address these research gaps. We use data from the National Comorbidity Survey Replication Adolescent Supplement (NCS-A), the largest U.S. nationally representative survey of adolescent mental health (N=10,123).

In Aim 1, we examine urbanicity as a source of effect heterogeneity that could contribute to inconsistent results. We combine propensity score subclassification with a survey design-based, weighted analysis to address confounding and positivity violations while maintaining national representativeness. We find evidence of effect heterogeneity; living in a disadvantaged neighborhood is associated with adolescent depression/anxiety if the neighborhood is in an urban center, but not if it is in the suburbs or rural area.

In Aim 2, we estimate the association between neighborhood disadvantage and cortisol in a subsample. We use propensity score methods coupled with regression to address non-random selection of families into neighborhoods and cortisol variability.

We find evidence of a heightened, yet resilient, cortisol response to a novel interview situation among adolescents in disadvantaged neighborhoods, thereby bolstering the evidence base suggesting that place may influence the stress response system.

In Aim 3, we address a question raised in Aim 2. Cortisol data are only available for a subsample of the NCS-A, so the estimated associations may be different from those estimated in the complete, nationally representative sample. This is a problem of generalizing results to a target population in the presence of non-random treatment assignment and a non-random two-stage selection mechanism. In Aim 3, we evaluate methods for generalizing such results using simulation and provide a tutorial for implementation.

This dissertation adds to the growing evidence suggesting that neighborhoods influence adolescent stress and mental health. We demonstrate the possible role of effect heterogeneity in explaining inconsistent findings. We show how causal inference methods can be used to address challenges in both neighborhood research and in large-scale studies involving biomarkers.

Thesis Readers

Gary Wand, M.D., Committee Chair
Professor, Department of Medicine

Thomas Glass, Ph.D., Advisor
Professor, Department of Epidemiology

Elizabeth Stuart, Ph.D., Advisor
Associate Professor, Department of Mental Health

Rosa Crum, M.D., M.H.S.
Professor, Department of Epidemiology

Acknowledgments

I am deeply appreciative of the support, encouragement, and guidance of others in completing this dissertation. In particular, I would like to thank my advisors, Drs. Tom Glass and Liz Stuart. Tom provided me with encouragement to tackle the issues that inspired me from the beginning of the Ph.D. program. His careful thinking and embracing of complexity modeled patience and thoroughness in approaching research questions. Liz provided me with innumerable hours of advice in applying causal inference methods. She pushed me further than I would have pushed myself in applying more advanced methods and becoming a better programmer (and frequent cluster user). In the process, I discovered that this meld of causal inference and social epidemiology was the research area that most excited me.

I would also like to thank my committee members, Drs. Rosa Crum and Gary Wand, as well as Drs. Kathleen Merikangas, Bill Eaton, Michael Rosenblum, and Iván Díaz for their many conversations and guidance over the years and iterations of this dissertation. They have deepened my understanding of both substantive and methodological issues within this dissertation and helped place these issues within the broader literature. This dissertation is truly a collaborative effort with all of them.

I've been fortunate during my time in the Ph.D. program to be the recipient of several generous funding mechanisms, including an NIMH pre-doctoral training

fellowship and Sommer scholarship. These have allowed me to dedicate more attention and time to research and less to worrying about paying the bills. Finally, I would like to thank my husband, William, for his support, grocery shopping, and whistling; my parents, Lisa and Kurt Rudolph, and brother Scott Rudolph, for their encouragement; and many friends for their enthusiasm.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Disease burden of anxiety and depression

A nationally representative survey of U.S. adolescents estimated that 11.7% of adolescents met criteria for ever having major depressive disorder or dysthymia and 31.9% have had an anxiety disorder at some point in their lifetimes. [91] Children with these disorders are more likely to miss school and sociodevelopmental opportunities, which can lead to difficulties in academic achievement and future employment, earnings, and family life. [11, 37, 65, 66] Moreover, these disorders often extend into adulthood where the economic costs are high. [96] Depressive disorders result in more than 500 million disability days per year in the U.S. and anxiety disorders result in more than 700 million disability days per year. [89] For major depressive disorder alone, the annual cost of lost productivity in the U.S. is more than 36 billion. [61]

Anxiety and depression are differentially distributed across countries with the U.S. having the highest prevalence of these disorders. [60] These disorders are also differentially distributed across smaller spatial scales, like region and neighborhood. In the following section, I will discuss research on how neighborhood factors may influence risk of anxiety and depression.

1.2 Neighborhood context and anxiety and depression.

1.2.1 Background

Research into the link between neighborhood context and mental health reaches as far back as Faris and Dunham's identification of the neighborhood patterning of schizophrenia in the 1930s. [5] Research into neighborhood-level risk factors—and social epidemiologic research in general—experienced a rebirth in recent decades, [25] but the field largely overlooked mental health outcomes. [93] In 2000, Mutaner et al argued that while research on physical health increasingly looked toward sociology, mental health research experienced a "rapprochement with biology and a retreat from sociological questions." [93] Since then, and in the past several years in particular, research into the relationship between neighborhood and mental health has surged.

To summarize the current state of this rapidly moving area of research, I conducted a literature review of studies conducted in high-income counties that examined the association between neighborhood-level characteristics and the mental health of children, adolescents, and adults. For literature published prior to 2008, I relied heavily on several review articles. [25,32,67,76,81] For literature published in 2008 and later, I systematically searched the Web of Science database using the search terms "mental health" and "neighborhood". I included mental health outcomes of depression, anxiety, and general psychological distress.

Most of the research used neighborhood SES as the neighborhood exposure. Other popular neighborhood-level exposures included collective efficacy (i.e., mutual trust among residents and their willingness to "intervene for the common goodâĂİ [120]),

disorder (i.e., âĂIJvisible cues indicating a lack of order or social control in the community" [112]), physical decay (e.g., abandoned, burnt-out, or deteriorating buildings [34]), and violence [32]. These neighborhood-level risk factors may be potential (and potentially competing) hypotheses about what it is about neighborhoods that confers additional risk of depression and anxiety. For this dissertation, I focus on neighborhood SES, which serves as a "proxy for a variety of specific features of neighborhoods potentially relevant to health." [24]

The evidence for an association between neighborhood SES and mental health is mixed. I reviewed 40 studies that examined this independent association after controlling for hypothesized individual-level confounding variables like age, sex, and individual-level socioeconomic status. Among adults, nine studies reported an inverse association between neighborhood SES and depression or depressive symptoms [4, 21, 33, 68, 72, 83, 97, 129] and eleven reported no association. [3, 22, 46, 52, 79, 124, 135, 140, 142, 146, 150] Three studies reported no association between neighborhood SES and anxiety. [27, 78, 142] Seven studies reported an association with psychological distress, [35, 70, 77, 114, 131, 132, 141] and three did not. [105, 143, 144] In studies of children and adolescents, four studies found an independent association between neighborhood SES and depression and/or anxiety [9, 70, 77, 149] and four did not. [15, 39, 130, 137]

Several of the referenced articles derived from key neighborhood studies: the Moving to Opportunity for Fair Housing Demonstration (MTO) and the Yonkers Project. These studies were designed for the purpose of studying the health effects of neighborhoods, and so likely contain enough neighborhood variability and within- and between-neighborhood sampling power to detect neighborhood-level effects. [24, 76]

Moving to Opportunity was a randomized control trial funded by the U.S. Department of Housing and Urban Development to evaluate the effects of moving public

housing residents to private housing in low-poverty, low-minority neighborhoods. [77] 4,248 families with children, residing in public housing in five cities—Baltimore, Boston, Chicago, Los Angeles, and New York—were randomized to one of three conditions: (1) receipt of vouchers and moving assistance to move to a low-poverty neighborhood (defined as less than 10% of residences at or below the federal poverty line); (2) receipt of vouchers but no moving assistance and no requirement about where to move; and (3) control group (no intervention). Randomization occurred between 1994 and 1997. This study's strengths lie not only in its purposeful neighborhood design, but also in its experimental set up. Because families are essentially randomized to poor and non-poor neighborhoods, this study was designed to be unaffected by unobserved confounding. Leventhal and Brooks-Gunn found that parents who were randomized to move to low-poverty neighborhoods in the New York City study site reported less distress two years later and male adolescents reported fewer anxiety and depressive symptoms. [77] No mental health benefits were found for female adolescents. Similarly, in analysis of five-year follow-up data from all five MTO sites, Kling et al found an association between randomization to the voucher group and lower psychological distress among adults. [70] In fact, the reduction in risk of serious mental illness among adults who moved was "comparable to that found in some of the most effective clinical and pharmacologic mental health interventions." [70] Kling et al also found mental health benefits for adolescents. However, unlike the analysis specific to the New York City study site, female—not male—adolescents randomized to the voucher group had lower psychological distress at follow-up.

The Yonkers Project was the result a 1985 federal court order to "remedy" the concentration of public housing in Yonkers, New York. [76] It is similar to MTO in that it was designed to look at the effect of housing desegregation. However, whereas

MTO was an experimental study due to randomization of the voucher assignment, the Yonkers Project was quasi-experimental. Instead of random assignment to treatment or control conditions, families in public housing signed up to receive vouchers to help them move to middle-income neighborhoods between late 1992 and 1994. The treatment group was composed of 173 Black and Latino families who signed up and were selected and the control group was composed of 142 families who signed up but were not selected. [27] In contrast to the associations found in the MTO study between being randomized to move to a low-poverty neighborhood and better mental health, the Yonkers Project found no evidence of such an association among the adults enrolled in the study. [27] However, adults who moved reported more positive perceptions of the neighborhood, including more general satisfaction with neighborhood conditions, lower perceived disorder, and lower violence exposure. In addition, adults who moved were more likely to be employed either part- or full-time than adults who did not move, controlling for sociodemographic characteristics.

1.2.2 Gaps

Despite the numerous studies on the relationship between neighborhood SES and anxiety and depression, the results are inconsistent. There could be at least several reasons for this. First, some inferences (either null or non-null) could be incorrect due to bias, possibly due to residual confounding. Even after adjusting for confounders, families that live in poor neighborhoods may not be exchangeable in terms of various sociodemographics with families that live in non-poor neighborhoods. Second, null inferences could be incorrect because wide confidence intervals may obscure true associations in some analyses. Third, inferences could be correct but inconsistent due

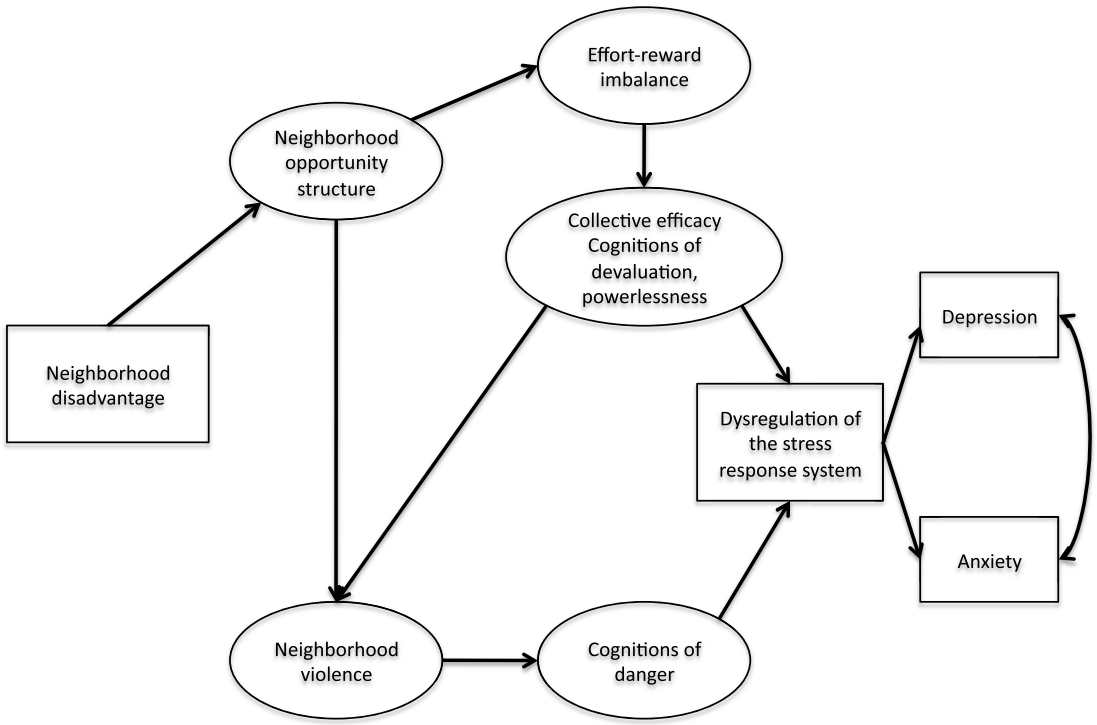
to effect heterogeneity. That is, there may be an association between living in a poor neighborhood and increased risk of depression or anxiety in some sub-populations but not in others. For example, level of urbanicity may identify sub-populations with differing strengths of association between neighborhood SES and depression/anxiety, thereby making urbanicity a modifer of the effect. The previous studies that found an association between neighborhood disadvantage and emotional disorders tended to sample from urban populations (e.g., [70, 77, 149]), whereas studies reporting no association tended to sample from non-urban populations and broader metropolitan areas (e.g., [39, 130]). We discuss how causal inference methods may be used to address these gaps in Section 1.5.

1.3 How neighborhood disadvantage may increase risk of depression and anxiety

As stated in Section 1.2.1, neighborhood SES is not, itself, a risk factor for poor mental health, but is a "proxy for a variety of specific features of neighborhoods potentially relevant to health. [And,] [u]npacking these specific features of neighborhoods and testing hypotheses about their relationship to specific health outcomes is what is needed to draw inferences regarding causal neighborhood health effects." [24] Research has not yet established which features of neighborhoods are risk factors for anxiety and depression, but neighborhood collective efficacy and violence currently have the strongest evidence bases. [2,17,31,41,44,71,79,80,92,125,135,149] A conceptual model of how neighborhood disadvantage may give rise to specific neighborhood features that increase risk of anxiety and depression is informed by Conditions-Cognitions-

Emotions theory [113] and is depicted in Figure 1.1.

Figure 1.1: Conceptual model. Measured variables are shown in rectangles and unmeasured variables are shown in ovals.



Historical segregation and redlining practices in the U.S. have resulted in persistent economic and racial homogeneity in many neighborhoods. Consequently, disadvantaged neighborhoods frequently lack what sociologist William Julius Wilson calls “neighborhood opportunity structure." [148] These neighborhoods “offer few legitimate employment opportunities, inadequate job information networks, and poor schools" and residents “lack contact or sustained interaction" with local institutions, individuals in the formal labor market, and individuals with some college education. [148]

Wilson developed his theory about neighborhood opportunity structure based

on qualitative research in urban neighborhoods, but quantitative research also supports this theory. Several experimental and quasi-experimental studies examined the employment effects of moving from high- to low-poverty neighborhoods. Based on Wilson's construct of neighborhood opportunity structure, we would predict that residents of high-poverty neighborhoods are more likely to be isolated from formal institutions, from the formal labor market, and from individuals well-connected with mainstream society. As such, we would predict that moving from high- to low-poverty neighborhoods would result in more employment opportunities. In an analysis of data from the Gatreaux Project in Chicago, Popkin et al found that low-income women who moved to middle-income neighborhoods were 13% more likely to be employed after moving, controlling for possible confounding factors like sociodemographic variables and work history. (This association was statistically significant.) [102] Similarly, analysis of the Yonkers Project data also found statistically significant associations between moving to higher-income neighborhoods and being employed. At two follow-up time points, two and seven years post-intervention, Fauth et al found that adults who moved were more likely be employed part- or full-time and less likely to be on welfare than those who did not move. [27, 28] However, analysis of the Moving to Opportunity data at both two and five years post-intervention found no statistically significant association between moving to middle-income neighborhoods and employment. [58, 70]

1.3.1 Neighborhood disadvantage to depression

Regarding the upper portion of the conceptual model in Figure 1.1, a lack of neighborhood opportunity structure may create an effort-reward imbalance (a situa-

tion that has received much attention in social epidemiologic research) that impedes collective efficacy on the neighborhood level and diminishes self-esteem and engenders feelings of powerlessness on the individual level. According to Wilson, legitimate efforts by residents of low-opportunity neighborhoods to improve their situation (*effort*) are likely to be ineffective (*lack of reward*), because of the "systematic blockage of opportunities in the environment of the inner city and society as a whole." [148] Research suggests that an effort-reward imbalance results in individuals feeling powerless, with little motivation and self-esteem, [128] and is cross-sectionally and longitudinally associated with physiologic stress [14, 56, 139] and mental disorder (particularly depression). [8, 38, 53, 73, 87, 88, 133, 139, 145] Although the vast majority of this research has been done in the context of the workplace, it is plausible that an effort-reward imbalance in a neighborhood context could also influence allostatic load and risk of poor mental health.

Perceived powerlessness and the lack of motivation wrought by an effort-reward imbalance "stymies collective efficacy," according to Sampson, Raudenbush, and Earls. [120] In fact, in describing their conceptualization of collective efficacy, Sampson et al implicate segregation and isolation, aspects of Wilson's neighborhood opportunity structure construct, as key influences, saying: "Economic stratification by race and place thus fuels the neighborhood concentration of cumulative forms of disadvantage, intensifying the social isolation of lower income, minority, and single-parent residents from key resources supporting collective social control." [120]

In addition, disadvantaged neighborhoods are likely to experience financial disinvestment and, because they typically lack the collective efficacy to advocate for public and private resources, this disinvestment may result in physical decay, including deteriorating and vacant buildings and dirty streets. These conditions may be

visual signals to residents that “no one cares," [147] leading to resident feelings that their neighborhood is devalued by the city government and those in the dominant society. [104] Persistent cognitions of devaluation may have the effect of lowering self-esteem and self-worth, [111, 136] qualities that are predictive of dysregulation of the stress response system, [69, 103] reduced hippocampal volume, [103] and increased risk of depression. [12, 13, 18, 119]

1.3.2 Neighborhood disadvantage to anxiety

Regarding the lower portion of the conceptual model in Figure 1.1, there are several ways in which neighborhood disadvantage could heighten the risk of experiencing cognitions of danger. As discussed in Section 1.3.1, a limited neighborhood opportunity structure may thwart collective efficacy, which has been shown to be associated with increased community violence. [120] In addition, because the opportunity structure has broken down in these neighborhoods, more adults are likely to be unemployed and/or out of the labor force, which increases the “likelihood that the residents will rely on illegitimate sources of income." [148] Because these illegitimate methods may be a target for police, this may also act to increase neighborhood violence.

More violent neighborhoods may increase feelings of ambient threat among neighborhood residents, because one is more likely to witness or hear about violence. Mirowsky and Ross write that although individuals living in these types of neighborhoods are at greater risk of victimization, the “likelihood of personal criminal victimization is low. On the other hand, residents live every day with the threat of victimization." [113] This ambient threat may increase psychosocial stress (“a heightened state of vigilance, alarm, or threat"), which may entail repeated or prolonged

activation of the stress response system. [75, 85] Do et al and Nazmi et al provide preliminary evidence of associations between neighborhood violence and markers of dysregulation of the stress response system. [26, 94] In addition, Findlay-Jones and Brown found these cognitions of danger to be predictive of anxiety disorders. [30]

1.4 Evidence of an association between neighborhood SES and allostatic load.

1.4.1 Background

We hypothesize that if there is an association between neighborhood SES and mental health, it operates through dysregulation of the stress response system. The beneficial form of the stress response system is called “allostasis," which McEwen and Seeman describe as the “process of maintaining stability by active means, namely, by putting out stress hormones and other mediators." [85] Allostasis allows for the “flight-or-fight" response that promotes survival in dangerous or predatory situations, but can also be elevated to facilitate adaptation to more mundane situations such as public speaking, interpersonal conflict, and even the simple act of waking up. [84] However, this same system may prove deleterious under repeated or prolonged activation—a concept that has been referred to as both “allostatic load" [85] and “weathering." [36] McEwen describes allostatic load as the “wear and tear on the body and brain caused by use of allostasis, particularly when the mediators are dysregulated, i.e., not turned off when stress is over or not turned on adequately when they are needed." [84] Allostatic load has been associated with impaired cardiovascular, immune, and

cognitive functioning. [84]

Several recent studies have found evidence of a relationship between neighborhood SES and allostatic load. In a paper published in 2010, Bird et al report an association between lower neighborhood SES and elevated allostatic load in a nationally representative sample (NHANES III) of over 13,000 adults from nearly 2,000 Census tracts. [7] To measure allostatic load, Bird et al used a nine-indicator summary measure (developed by Seeman et al for the NHANES sample) to capture "cumulative physiological dysregulations across multiple physiologic regulatory systems [e.g., cardiovascular, inflammatory, and metabolic]." [126] Another recent study from the Multi-Ethnic Study of Atherosclerosis (MESA), a ten-year cohort study of adults aged 45-84 in six cities in the U.S., found evidence of an association between neighborhood SES and two inflammatory markers of allostatic load. [94] Nazmi et al assessed cross-sectional and longitudinal associations between three inflammatory markers (fibrogen, C-reactive protein, and interleukin-6) and neighborhood SES among all 5,370 MESA cohort members. These authors found neither a cross-sectional nor longitudinal association between C-reactive protein and neighborhood SES. However, they did find a cross-sectional association between neighborhood SES and fibrogen and a longitudinal association between neighborhood SES and interleukin-6. [94] Similarly, Pollitt et al and Petersen et al found evidence of independent associations between inflammatory markers and neighborhood SES among adults in a multi-site study in the U.S. and residents of southwestern Pennsylvania, respectively. [99, 101]

Although cortisol is only one component of the complex, multi-faceted stress response system, [85] it is an important indicator of hypothalamic pituitary-adrenal (HPA) axis reactivity. [84] A low locus of control is a significant predictor of cortisol response in humans [23, 103] and primates [121], low-ranking primates show consis-

tently elevated HPA activation, [122] and humans who are lonely show higher cortisol levels upon waking. [134] Studies on rodent pups have shown that early life psychosocial stress (e.g., separation from mothers, poor maternal behavior) has lasting effects on HPA axis reactivity, and thus, cortisol levels. [86, 100] And researchers have found that children from high-risk families show dysregulation of the HPA axis lasting well into adulthood. [106]

In addition, cortisol dysregulation has been shown to have multiple effects on the body. Cushing's Disease—a disorder caused by a tumor on the pituitary glad resulting in high levels of cortisol—exemplifies some of these effects. These include cognitive effects such as increased risk of major depression and anxiety disorders, metabolic effects such as central obesity and diabetes, immune effects such as increased risk for infection, and cardiovascular effects such as hypertension and heart disease. [84] The central role of cortisol is underscored by the fact that Cushing's disease is effectively treated by removing the tumor, which allows cortisol levels to return to normal. [98]

Adverse conditions in neighborhood and family environments have been linked to both cortisol levels and cortisol reactivity, although the evidence is mixed. In adults, some studies have yielded associations between neighborhood- and individual-level low socioeconomic status (SES) and cortisol diurnal levels—specifically lower waking levels, [43] higher average levels, [19, 20] and less steep declines over the course of the day [1, 26, 43, 57] though others have found null or opposite results. [1, 19, 43] Do et al also found that neighborhood violence was associated with lower cortisol levels at awakening and less steep initial declines. [26] In children, studies have reported associations between individual-level disadvantage (including low SES, exposure to stressful life events, and family adversity) and lower morning cortisol levels, [6, 106] higher average cortisol levels, [6, 29, 59, 106] and less steep declines. [40, 59, 82] In addi-

tion, some have suggested a curvilinear (upside-down u-shaped) association; children and adolescents exposed to the most stressful conditions have cortisol levels that resemble those of non-disadvantaged individuals, possibly due to eventual blunting of the HPA axis after repeated activations. [6, 40]

1.4.2 Gaps

The evidence for an independent association between adverse neighborhood conditions and salivary cortisol in adolescents is extremely limited. Studies conducted to date provide preliminary evidence that neighborhood disadvantage is associated with higher average resting cortisol levels [10, 16] and greater cortisol reactivity. [42] However, these studies have been observational and have used standard regression adjustment techniques to address non-random selection of families into neighborhoods without first examining whether or not adolescents living in different kinds of neighborhoods are comparable or whether regression adjustment would instead have to rely on extrapolation. In addition, the studies have been based on small, racially homogeneous samples in single urban areas, which could compromise generalizability in the presence of treatment effect heterogeneity. [10, 16, 42]

1.5 Causal inference methods to address gaps

Causal inference methods may be used to address gaps in (1) research examining the association between neighborhood SES and anxiety and depression and (2) research in examining the association between neighborhood SES and cortisol. As argued by Rubin [115] and Hernan [47], when randomized trials are not possible (as

is frequently the case when studying neighborhood residence), using causal inference methods in the design and analysis of observational studies is the next best option. Whether or not causal inference is possible in such studies is open for debate. [95] However, the casual inference framework and methods can nevertheless be used to design an analysis that carefully assesses and evaluates assumptions and utilizes methods to minimize reliance on possibly untenable assumptions.

The Neyman-Rubin causal model [127] is a framework for causal inference in observational settings. Using this model, the researcher explicitly states the assumptions that must hold to make a statement about a causal effect. The researcher then provides a logical argument as to the extent to which each of the assumptions hold, provides data in support of the extent to which each assumption holds, and/or conducts sensitivity analyses to violations of the assumptions. [115] The central assumptions for causal inference include: (1) The stable unit treatment value assumption (SUTVA), which Rubin defines as "the value of Y for unit u when exposed to treatment t will be the same no matter what mechanism is used to assign treatment t to unit u and no matter what treatment the other units receive." [117] In other words, SUTVA has two parts. First, it assumes that there is only one version of each treatment. Second, it assumes that the treatment that one person receives does not influence the effect of the treatment that another individual receives. This is also referred to as assuming no interference. In the context of neighborhood research, the first part of SUTVA requires that the neighborhood-level exposure is the same for all participants in the study. The second part of SUTVA assumes that the neighborhood exposure for one participant does not affect the potential outcomes of another participant. This second assumption is problematic in most neighborhood studies where many participants are clustered within a single neighborhood. However, it is less of a concern for

this dissertation work, as the clustering of participants within neighborhood is very low. We further discuss possible violations of SUTVA in Section 5.3.1.2. (2) Each unit must have a non-zero probability of being exposed to any one of the causes/ treatments. [50] This is also called the positivity assumption and means that exposure to a cause/ treatment cannot be deterministic—there must be the potential for randomness in treatment assignment. (3) Treatment assignment is random, possibly conditional on background variables. [50] This is also known as the independence assumption or exchangeability. (4) Sampling of units from the population of units, U, is random, possibly conditional on background variables. [115] This allows that "All probabilities, distributions, and expected values involving variables are computed over U." [50]

Much work has gone into developing methods that can provide a "statistical solution" [50] to the fundamental problem of causal inference, which is that one cannot both simultaneously observe alternative treatment/ causal states. For this dissertation, I will discuss methods that address the assumptions of positivity, independence in treatment assignment, and independence in sampling. Imai et al decomposed estimation error as

$$Error = (e_{\Delta x} + e_{\Delta u}) + (e_{Ax} + e_{Au}), \tag{1.1}$$

where the subscript x denotes observed error and the subscript u denotes unobserved error, A is exposure assignment, and Δ is sample assignment. [54] In observational studies, non-random treatment assignment results in e_A, also called confounding bias. In both observational and experimental studies, non-random sample assignment results in e_Δ, also called lack of generalizability or non-transportability. Non-random sample assignment can be thought of more generally as a missing data problem that

applies to sample selection, right-censoring (e.g., loss to follow-up), and non-response. Such non-random assignment is particularly problematic in the presence of treatment effect heterogeneity. Without treatment effect heterogeneity, the treatment effect is assumed to be constant for everyone, and thus, the marginal effect would not be affected by non-random selection. In the real world, however, it is unlikely that a particular treatment or exposure would have the same effect on everyone. In the presence of effect heterogeneity, non-representative missing data in the form of sample selection, censoring, or non-response that depends on effect modifiers may result in biased marginal effect estimates as well as biased conditional effect estimates if all sources of heterogeneity are not accounted for. Addressing both types of bias becomes more tractable if we can assume independence in sampling and treatment assignment conditional on background variables. This assumes that $e_{\Delta u}$ and e_{Au} in Equation 1.1 are negligible. Causal inference methods that seek to achieve conditional independence in treatment and selection assignment include weighting, [51], propensity score methods, [110] G-computation, [107] and numerous double-robust estimators such as augmented inverse probability weighting (AIPW) [108] and targeted maximum likelihood estimation (TMLE) [138]. Applying the Neyman-Rubin Causal Model to the research gaps identified in Sections 1.2.2 and 1.4.2, suggests several causal inference methods that may be used to address non-random selection of families into neighborhoods and non-random selection of adolescents into the study sample.

1.5.1 Methods to address non-random assignment of families into neighborhoods

The fact that neighborhood is not randomly assigned threatens the independence of treatment assignment assumption as well as the positivity assumption. Combined, these assumptions can be called the strongly ignorable treatment assignment assumption, which says that responses to a treatment are independent of any covariates, and that the probability of receiving a treatment conditional on those same covariates is greater than 0 and less than 1. [110] Propensity scores are one way to reduce bias due to violation of this assumption.

A propensity score is defined as the probability of treatment assignment conditional on a vector of observed variables. [110] It is a type of balancing score (in fact, it is the coarsest balancing score) because exact matching on propensity score ensures that the treatment groups being compared are balanced across a vector of covariates. Rosenbaum and Rubin proved the following theorem: "At any value of a balancing score, the difference between the treatment and control means is an unbiased estimate of the average treatment effect at that value of the balancing score if treatment assignment is strongly ignorable," which holds under both large and small sample theory. [110] They then showed through corollaries that pair matching on the propensity score, subclassification based on the propensity score, and covariate adjustment on the propensity score all produce unbiased estimates of treatment effects where treatment assignment is strongly ignorable. However, exact matching on the propensity score is typically impossible. Rosenbaum and Rubin proved that approximate propensity score matching, subclassification, and adjustment reduce bias [110] and Rubin showed through simulation that this bias reduction is substantial. [116]

Variables in a propensity score equation should not be affected by the exposure. In other words, they should not be mediators of the exposure-outcome relationship. Opportunities and constraints associated with neighborhood disadvantage may influence several variables included in the propensity score model such as household income, family structure, and even maternal education and maternal age at child's birth when considered as part of a multigenerational feedback loop. Such variables may have both mediating and confounding roles. To the extent to which they act as mediators, including such variables in the propensity score model will remove the indirect effect of neighborhood disadvantage on mental health that operates through these variables. However, as they may also contribute to confounding, failure to control for them may bias the results. In this dissertation, we include several of the variables that may be less influenced by current neighborhood residence (e.g., maternal education, maternal age at birth of the child) in the propensity score model to control for family socioeconomic status. We control for other measures of family socioeconomic status in the regression analysis in separate models.

Thus, propensity score matching and subclassification are two causal inference methods that can reduce bias stemming from non-random assignment of families to neighborhoods. These methods have several advantages over traditional model-based regression adjustment. First, propensity score balance between the two groups can be compared—residual imbalance in the propensity scores is indicative of "potential bias in estimated treatment effects." [110] Second, propensity score matching/subclassification reduces variance in the ATE, because this technique makes the covariate distributions between the groups more similar. [110] Third, propensity score matching/subclassification makes subsequent model-based analyses more robust to model misspecification. [110] Fourth, propensity score matching/subclassification or adjust-

ment can offer a way to analyze a dataset that has too small of a sample size to support controlling for all necessary variables by model-based methods. [110]

Propensity score matching and subclassification be thought of as a data preprocessing technique to be done prior to and in conjunction with whatever outcome analysis was originally planned (e.g., regression) in order to best control for confounding. [49] Not only is propensity score matching/subclassification ideally combined with an outcome analysis like regression, multiple matching methods can be combined for further bias reductions. Rubin and Thomas demonstrated through analytic proofs and simulation that matching on the scalar propensity score could be combined with additional matching methods, such as exact matching and matching within calipers of the Mahalanobis distance, on a subset of highly influential covariates to reduce bias both from the linear relationship between the outcome and vector of covariates as well as from any non-linear relationships between the outcome and subset of particularly influential covariates included in the exact and Mahalanobis matching. [118]

Propensity score matching and subclassification methods can also be used to address violations of the positivity assumption. Violations of the positivity assumption may be structural. For example, in an observational study, a structural violation could occur if doctors only give Treatment A to patients in critical condition. [47] Practical violations of positivity occur if finite sample size prevents seeing individuals with the same vector of covariate values with different treatments. This is particularly relevant to neighborhood research. Given the history of racial segregation and redlining housing policies in the United States, in finite samples, there may be no poor, African Americans living in high-SES neighborhoods to compare with poor, African Americans living in low-SES neighborhoods. In this scenario, adjustment becomes problematic, because African Americans living in poor communities with a certain

vector of covariate values may not have exchangeable counterparts living in non-poor communities, thereby resulting in extrapolation in regression models. [48] Propensity score tools help by allowing the researcher to limit the analytic sample to the subset of individuals in one treatment group who have propensity scores that are approximately equal to propensity scores in the comparison treatment group. In other words, the analyst can limit to the subset to those whose propensity scores fall within the region of overlap (area of support) in order to maintain structural positivity. [45, 74] While restricting the sample to the area of overlap helps in meeting the positivity assumption, it comes at the price of reducing sample size and possibly compromising external validity. In scenarios where positivity problems are extensive—as may be the case in neighborhood research in segregated communities—much of the sample may be uncomparable.

1.5.2 Methods to address non-random selection of adolescents into the study sample

Non-random selection of participants into a study threatens the causal inference assumption of random sampling. If one also assumes a constant treatment effect and no unobserved confounding, then the methods described above can be used to address this source of bias. The vector of covariates would need to include effect modifiers that are predictive of sample selection in addition to confounders of the treatment-outcome relationship. However, controlling for sample selection is less straightforward in the presence of effect heterogeneity. Effect heterogeneity means that the effect of the treatment on the outcome differs based on the value or values of one or more covariates. For example, the effect of living in a poor neighborhood on risk of anxiety

may be greater for those living in urban areas than for those living in rural areas. Let's say that Sample A contains 90% urban residents and Sample B contains 10% urban residents. Even if urbanicity is included as a covariate in a regression analysis in Sample A, the conditional association between neighborhood and anxiety will still differ between Sample A and Sample B because of effect heterogeneity.

Effect heterogeneity may be integral to the research question or may be a nuisance. Estimating the conditional effects separately by value of the effect modifier may be desirable in scenarios where identification of specific differences in treatment effect are integral to the research question of interest—for example, if one were interested in an underlying mechanism or in identifying subgroups most likely to benefit from a treatment. [47] One can do this by including an interaction term between the effect modifier and treatment/exposure in a parametric regression model. This approach gives effect estimates that are conditional not only on a particular vector of covariates, but also on a particular value of the effect modifier. However, this approach is no longer practical when the number of effect modifiers becomes moderate to large, as there may not be enough data to estimate the interaction terms with meaningful precision.

Effect heterogeneity may also need to be addressed even in cases where it is viewed as a nuisance—for example, in research questions interested in average population effects, including whether an effect in one population can be applied to another. [47] When an average population effect and/or transportability are the primary goals, common methodological options include inverse probability weighting, standardizing to the target population [47] or imputing data for those who were not selected (however, this is typically not recommended since the number of individuals not selected may be much larger than the number of individuals selected and little to no data

may be available for those not selected). These methods will compute an average effect—conditional or marginal—standardized to the target population of interest. Moreover, they can be combined with methods to deal with confounding to address both sources of non-randomness. [55, 109, 123]

Addressing both sources of non-randomness is especially important for neighborhood research, because of both the non-random process by which families are selected into neighborhoods and the fact that neighborhood study samples are rarely nationally representative or population-based. However, there has been little to nothing written about methods—especially double robust methods—that address both sources of non-randomness in the neighborhood literature or in the broader epidemiologic literature.

1.6 Overview of study aims

As summarized in the previous sections, studies of the association between neighborhood SES and adolescent anxiety/depression and studies of the association between neighborhood SES and cortisol have generated inconsistent results. Contributing factors to these inconsistent findings may include non-transportability of effects estimates across study populations due to effect heterogeneity as well as incomplete control for confounding and structural positivity due to neighborhood selection/segregation. The goal of this dissertation is to apply several of the causal inference methods discussed in Section 1.5 to address these research gaps. We use data from the National Comorbidity Survey Replication Adolescent Supplement (NCS-A), the largest U.S. nationally representative survey of adolescent mental health (N=10,123), which was conducted between 2001 and 2004. [62–64, 90] This dissertation has three

aims.

1. In Aim 1, we examine one source of effect heterogeneity that could contribute to inconsistent results: urbanicity. Previous studies reporting an association between neighborhood disadvantage and depression/anxiety have been limited to urban populations, whereas studies reporting no association tended to sample from non-urban populations. We hypothesize that living in a disadvantaged neighborhood in an urban versus rural area likely entails exposure to a different set of stressors, such as crowding, lack of green space, and exposure to violence, which may differentially influence mental health. We combine propensity score subclassification with a survey design-based, weighted analysis to address confounding and structural positivity stemming from non-random assignment of neighborhood residence while maintaining the national representativeness of the sample. We also perform a sensitivity analysis for unobserved confounding.

2. In Aim 2, we estimate the association between neighborhood disadvantage and cortisol—a biomarker of the HPA axis. We examine the association in a large, diverse, population-based sample of adolescents, thereby addressing the gap of limited generalizability of previous research. We combine propensity score matching methods with regression adjustment to address non-random selection of families into neighborhoods. We also combine traditional propensity score matching with exact and caliper matching to strengthen bias reduction for variables shown to be influential drivers of cortisol variability (a key challenge in non-laboratory settings). We assess the sensitivity of our results to an unobserved confounder.

3. In Aim 3, we address a question raised in Aim 2. Cortisol data are only available

for a subsample of the NCS-A. Thus, the estimated associations may be different from those estimated in the complete, nationally representative sample because of effect heterogeneity. To estimate the average population effect of living in a disadvantaged neighborhood on cortisol, we would need to control for non-random assignment of families into neighborhoods as well as non-random selection into the cortisol sample. Put more broadly, this is a problem of generalizing results from a non-randomized study sample to a specified target population. In Aim 3 we use simulation to compare practical estimators for the population average treatment effect that simultaneously account for non-randomized treatment assignment and sub-sample selection from a population-based cohort, thereby simultaneously addressing internal and external validity. We also provide a tutorial for applied researchers on how to implement such methods.

This dissertation is important for four main reasons. First, we add to the growing body of evidence suggesting that neighborhoods influence adolescent stress and mental health. Second, we demonstrate the importance of effect heterogeneity and its possible role in explaining inconsistent findings. Third, we show how causal inference methods can be used to address challenges in both neighborhood research and in large-scale studies involving biomarkers. Fourth, we evaluate simple estimators of the marginal mean treatment effect that simultaneously account for non-random treatment assignment and sample selection and show how their application can lead to less biased, more transportable estimates that may be more easily interpreted and understood by policymakers.

Chapter 2

The influence of urbanicity on the association between neighborhood disadvantage and adolescent emotional disorder

This chapter's research has been published. [34] All code associated with this chapter can be found here: https://github.com/cherrygarcia/Aim1.

2.1 Abstract

Inconsistent evidence of a relationship between neighborhood disadvantage and adolescent mental health may be, in part, attributable to heterogeneity based on urban or rural residence. Using the largest nationally representative survey of U.S. adolescent mental health available, we estimated the association between neighborhood disadvantage and adolescent emotional disorders and the extent to which urbanicity modified this association. The National Comorbidity Survey Replication Adolescent Supplement (NCS-A) sampled adolescents aged 13-17 years (N=10,123). Households were geocoded to Census tracts. Using a propensity score approach that addresses bias from non-random selection of individuals into neighborhoods, logistic regression

models were used to estimate the relative odds of having a DSM-IV emotional disorder (any past-year anxiety disorder, major depressive disorder or dysthymia) comparing similar adolescents living in disadvantaged versus non-disadvantaged neighborhoods in urban center, urban fringe, and non-urban areas. The association between neighborhood disadvantage and emotional disorder was more than twice as large for adolescents living in urban centers versus non-urban areas. In urban centers, living in a disadvantaged neighborhood was associated with 59% (95% confidence interval: 25-103%) increased adjusted odds of emotional disorder. Urbanicity modifies the relationship between neighborhood disadvantage and emotional disorder in adolescents. This effect modification may explain why evidence of a relationship between neighborhood disadvantage and adolescent mental health has been inconsistent. Recognizing the joint influence of neighborhood socioeconomic context and urbanicity may improve specificity in identifying relevant neighborhood processes.

2.2 Introduction

A nationally representative survey of U.S. adolescents showed that nearly one-half met criteria for ever having a mental disorder and nearly one-fourth met criteria for ever having a mental disorder with severe impairment. [24] Children with mental disorders are more likely to miss school and sociodevelopmental opportunities, which can lead to difficulties in academic achievement and future employment, earnings, and family life. [3, 11, 17, 18] Moreover, these disorder often extend into adulthood. [27]

An array of factors at multiple levels—genetic, family, and environmental—may influence mental health. Adolescents may be particularly susceptible to the influences of their neighborhood environment, because of the shift from the home to the

external environment during this developmental period. [40] Neighborhood socioeconomic disadvantage (henceforth, neighborhood disadvantage)—a "proxy for a variety of specific features of neighborhoods potentially relevant to health," [31]—is the most widely studied neighborhood characteristic that may influence mental health.

Research into the relationship between neighborhood disadvantage and child/adolescent emotional disorders (i.e., anxiety and depressive disorders) has resulted in inconsistent evidence [e.g., [12, 19, 23, 35, 42]]. No study has explained these discrepant findings, nor identified which disadvantaged neighborhoods may be particularly detrimental to child/adolescent mental health. A characteristic that identifies neighborhoods that have systematically different relationships between neighborhood disadvantage and mental health—and therefore could offer an explanation for the inconsistencies—is called an effect modifier.

It is possible that urbanicity is an effect modifier of the neighborhood disadvantage-adolescent mental health association. The previous studies that found an association between neighborhood disadvantage and emotional disorders tended to sample from urban populations [e.g., [19, 23, 42]], whereas studies reporting no association tended to sample from non-urban populations and broader metropolitan areas [e.g., [12, 35]]. Living in a disadvantaged neighborhood in an urban area likely entails exposure to a different set of stressors than living in a disadvantaged neighborhood in a rural area. Although certain exposures detrimental to mental health are more prevalent in rural disadvantaged neighborhoods, such as a lack of access to mental health care and resources, [25] other stressors may be more prevalent in disadvantaged urban neighborhoods. Lack of green space, noise, residential instability and exposure to violence in the neighborhood and in the neighborhood school may be more prevalent in disadvantaged urban neighborhoods, and research has linked these stressors

to emotional disorders [e.g., [9, 21, 36]]. Therefore, we hypothesize that urbanicity modifies the association between living in a disadvantaged neighborhood; specifically, we hypothesize that the positive relationship between neighborhood disadvantage and emotional disorders is greater in urban areas than in non-urban areas.

A central challenge to testing this hypothesis, and to neighborhood research in general, is addressing non-random selection of families into neighborhoods. [31] The population of one neighborhood may look very different (e.g., in terms of race/ethnicity, income distribution) from another. This issue is typically addressed with regression adjustment. However, when trying to estimate the effect of living in one type of neighborhood versus another, this form of adjustment poses a problem if adolescents living in disadvantaged neighborhoods do not have similar (also called exchangeable) counterparts living in non-disadvantaged neighborhoods. [14] To address this challenge, researchers have increasingly drawn on propensity score tools. [31] However, we know of no study of the relationship between neighborhood and adolescent mental health that has used these methods.

Our objectives were to estimate the association between neighborhood disadvantage and emotional disorders in adolescents within levels of urbanicity and to test for effect modification by urbanicity, addressing non-random selection into neighborhoods through a propensity score approach. We used data from the National Comorbidity Survey Replication Adolescent Supplement (NCS-A), a nationally representative survey of adolescent mental health. Design-based, survey-weighted logistic regression models were combined across propensity score subclasses to estimate the relative odds of having an emotional disorder comparing similar adolescents living in disadvantaged versus non-disadvantaged neighborhoods in urban center, urban fringe, and non-urban areas.

2.3 Methods

The NCS-A is a nationally representative sample of adolescents in the continental United States designed as a survey of adolescent DSM-IV syndromes/disorders. The background, sampling and recruitment methods, and weighting scheme have been described elsewhere. [15, 16] Adolescents aged 13-17 years were sampled via dual-frame household and school samples. Professional interviewers at the Survey Research Center at the Institute for Social Research at the University of Michigan administered face-to-face, laptop computer-assisted personal interviews in the adolescentsâĂŹ homes between February 2001 and January 2004. While an adolescent was being interviewed, one parent or parent surrogate was asked to complete a self-administered questionnaire; those who did not complete this questionnaire were asked to complete a short-form version. Weighting procedures adjust the sample of 10,123 adolescents to be representative of students in the U.S. population in 2000 and are described in more detail elsewhere. [15] Informed assent and consent were obtained from each adolescent and his/her parent or guardian. The Human Subjects Committees of Harvard Medical School and the University of Michigan approved recruitment and consent procedures.

Participant residential addresses were geocoded to 3367 U.S. Census tracts by the Survey Research Center at the Institute for Social Research at the University of Michigan. Excluding adolescents residing in Census tracts with missing or inestimable Short Form 3 (SF3) indicators resulted in a final sample size of 10,074 adolescents in 3347 Census tracts.

2.3.1 Contextual measures

2.3.1.1 Neighborhood Disadvantage

Neighborhood socioeconomic status (SES), based on a summary score derived from factor analysis by Diez-Roux et al., [4] has been used in multiple epidemiologic studies [e.g., [32]]. It is composed of six indicators from the U.S. Census SF3: 1. log median household income, 2. percent households with interest, dividend, or rental income, 3. log median value of housing units, 4. percent persons over age 25 with high school degree, 5. percent persons over age 25 with college degree, and 6. percent persons in executive, managerial, or professional specialty occupations. Indicators were transformed to z-scores and summed to make a normally distributed summary score. We re-examined the factor structure and fit statistics in our sample and confirmed the one-factor structure. In our sample, the summary score had a Cronbach's alpha of 0.83 and ranged from -13.6 to 17.8 with a median value of -0.36. Neighborhoods in the lowest SES tertile were defined as disadvantaged, and neighborhoods in the two upper tertiles were defined as non-disadvantaged.

2.3.1.2 Urbanicity

Urbanicity was analyzed using the following Census-derived categories: (1) large-mid urban center (henceforth, urban center), (2) urban fringe, or (3) non-urban. Adolescents were categorized as living in a large-mid urban center area if they resided in a central county of a metropolitan statistical area. Urban fringe areas were defined as non-central counties of a metropolitan statistical area. Non-urban areas were defined as small to large towns or rural areas. The non-urban category was used as

the reference category for the two dummy variables of urban center and urban fringe.

2.3.2 Individual Measures.

2.3.2.1 Outcome measures

During face-to-face interviews, all NCS-A participants were administered a modified version of the World Health Organization Composite International Diagnostic Interview (CIDI) to assess the presence of mental health disorders/syndromes consistent with the DSM-IV. [16, 24] Past 12-month emotional disorder was defined as any anxiety disorder, major depressive disorder, or dysthymia. In accordance with prior recommendations, diagnostic algorithms for emotional disorders included only data obtained from the adolescent. [1]

2.3.2.2 Covariate measures

As recommended by Leventhal and Brooks Gunn, [22] covariates included adolescent age (in years), race/ethnicity, immigrant generation, household income, maternal age at birth, maternal level of education, family structure, and region of residence (i.e., Northeast, Midwest, South, West). Adolescent age, race/ethnicity, immigrant generation, and family structure were obtained during the adolescent interview. Race/ethnicity was categorized as non-Hispanic white, non-Hispanic black, Hispanic, and Other. Immigrant generation was categorized as first (foreign born), second, or third or greater. Family structure consisted of two variables assessing whether or not the adolescent had lived his/her whole life with his/her (1) mother and (2) father. Maternal education, maternal age at birth (in years, modeled with linear and

quadratic terms), and household income (log-transformed) were obtained from the parent questionnaire. Maternal education was divided into four ordered categories: less than high school, high school, some college, and college graduate.

2.3.3 Statistical Analysis

We used multiple imputation by chained equations to address missing data. [38] This procedure has been shown to be an effective way to address missing data in large datasets and requires less strict assumptions than excluding those with missing data. [37] Variables in the imputation model included variables hypothesized to influence response as well as all variables to be used in the analysis.

In implementing the propensity score approach, we followed previous recommendations for combining propensity score subclassification with complex survey data. [43] A propensity score [30] was estimated for each adolescent that is the predicted probability of living in a disadvantaged neighborhood at the time of the study as a function of the covariates specified above and their interactions with sex. It was estimated using an unweighted logistic regression model, as this is what is recommended by Zanutto [43], the rationale being that the propensity score model is used for matching purposes only and not to make population-level inferences. Variables in a propensity score equation should not be affected by the exposure. In other words, they should not be mediators of the exposure-outcome relationship. Opportunities and constraints associated with neighborhood disadvantage may influence several variables included in the propensity score model such as household income, family structure, and even maternal education and age when considered as part of a multigenerational feedback loop. However, we believe that it is essential to control for these covariates because

of their role as potential confounders.

To ensure that we compared participants living in disadvantaged neighborhoods to exchangeable participants living in non-disadvantaged neighborhoods (formally known as structural positivity), we restricted the analysis to the region with overlapping propensity scores of individuals living in disadvantaged and non-disadvantaged neighborhoods (see Figure 2.1 below). This resulted in the exclusion of 474 adolescents (4.7%) who had a propensity score outside the region of overlap in any of the 100 imputed datasets for a total sample size of 9,600 adolescents.

Propensity score subclassification controlled for the bulk of confounding. Figure 2.2, below, compares the imbalance in covariates (as measured by standardized mean difference) before and after subclassification. Covariate balance between the two exposure groups was achieved (all standardized mean differences less than 10%) with the nine subclasses. We used nine subclasses, because this was the largest number of subclasses that for which there were at least 20 participants in each of the two exposure groups.

Figure 2.1: Propensity score distributions comparing those in the disadvantaged neighborhood group to those in the non-disadvantaged neighborhood group.

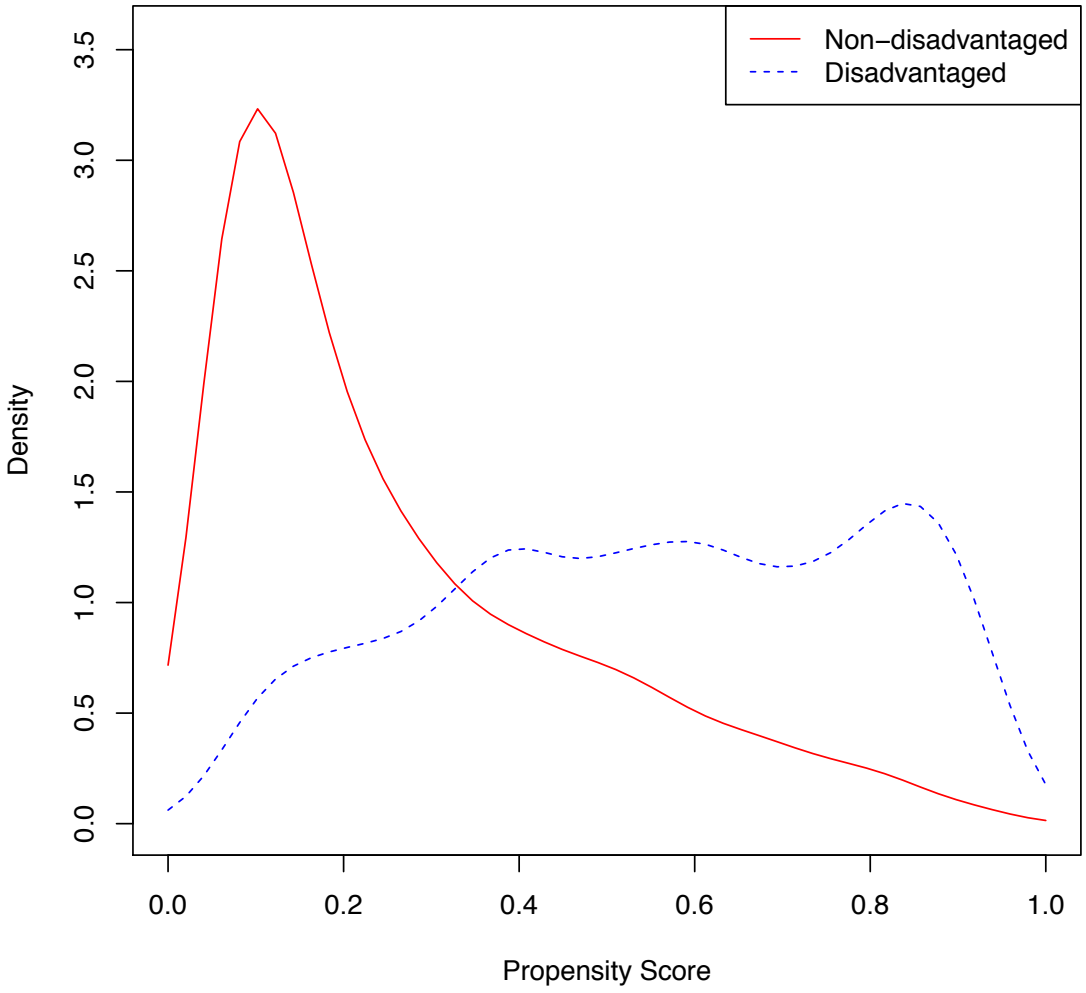
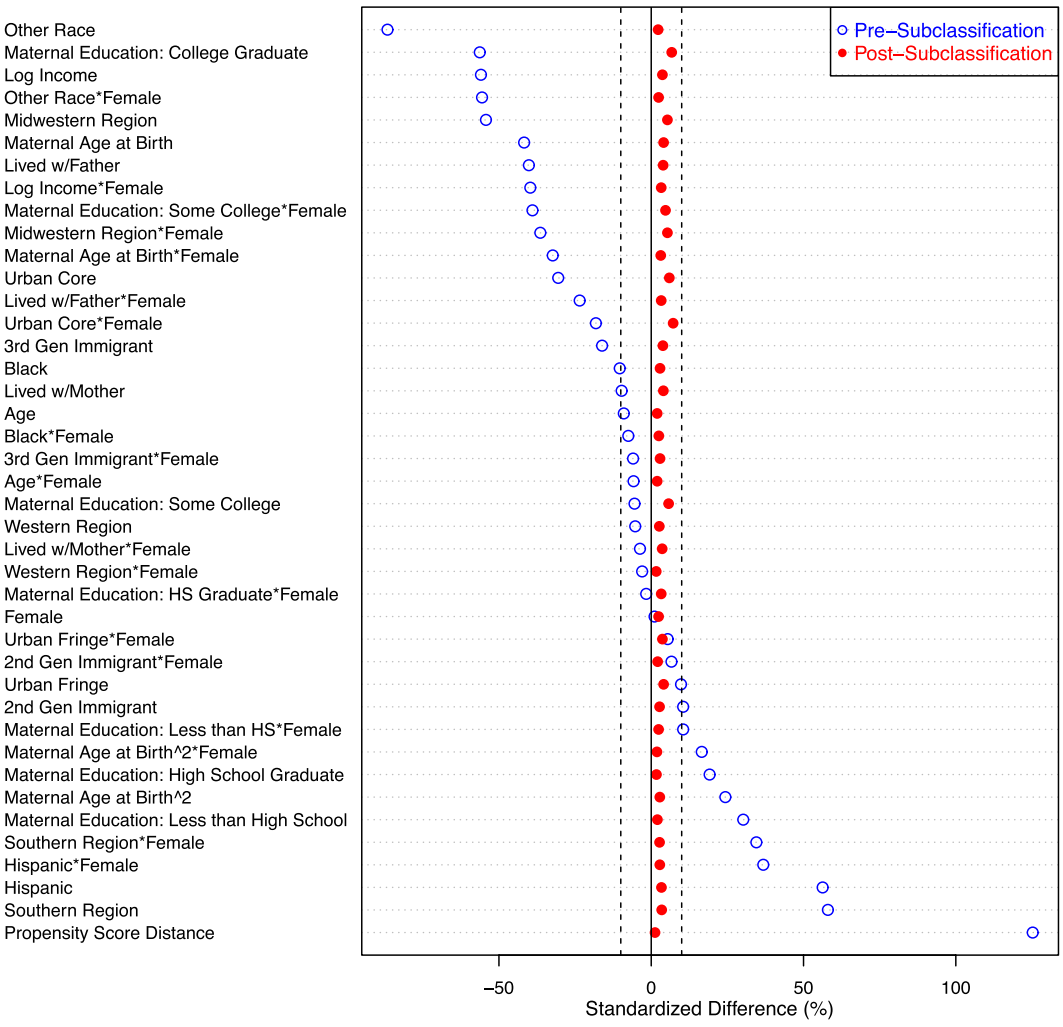


Figure 2.2: Covariate balance pre- and post-subclassification. Plotted points represent the standardized mean differences (difference in means between the disadvantaged neighborhood group and non-disadvantaged neighborhood group standardized by the standard deviation in the disadvantaged group) for each covariate. Open dots represent standardized mean differences in the pre-propensity score subclassification data. Closed dots represent standardized mean differences in the post-propensity score subclassification data.



We ran design-based, survey-weighted logistic regression models within each propensity score subclass to estimate the log odds of having an emotional disorder as a function of neighborhood disadvantage, neighborhood disadvantage × urbanicity interaction (to assess effect modification), and propensity score (to help control for

residual confounding). Variances were approximated using Taylor linearization using the survey package in the R statistical language (version 2.14.1). Coefficient estimates and variances were combined across subclasses using the Mantel-Haenszel method. This method assumes that each subclass is estimating one, common effect. An advantage of this method over weighting by the total weights in each subclass is that it incorporates other survey design features, such as sampling strata. In addition, we found that resulted in smaller variances. Then, these average effects were pooled across the 100 imputed datasets using Rubin's combining rules. [33] All analyses were performed using R.

2.4 Results

Table 2.1 presents design-based, weighted descriptive statistics of the sample by neighborhood disadvantage status. The mean age of the participants was just over 15 years, and 49% were female. A greater percentage of adolescents in disadvantaged neighborhoods were Hispanic or Black. Similar percentages reported living with their mother for their whole life but fewer adolescents in disadvantaged neighborhoods had lived with their father for their whole life. In non-disadvantaged neighborhoods, mean household income was about 30 000 dollars greater, and more mothers had a college education and were about two years older at the time of birth. Adolescents living in disadvantaged neighborhoods were more likely to live in the urban fringe and non-urban areas. The prevalence of emotional disorders was slightly higher in disadvantaged neighborhoods.

Table 2.1: Design-based, weighted NCS-A sample characteristics by neighborhood disadvantage status.

	Disadvantaged (n=3,597)			Non-disadvantaged (n=6,003)		
	Mean	95% CI		Mean	95% CI	
Female	0.49	0.46	0.53	0.49	0.47	0.51
Age	15.17	15.02	15.31	15.19	15.05	15.34
Race/ethnicity						
Hispanic	0.24	0.19	0.29	0.11	0.09	0.13
Black	0.28	0.23	0.32	0.09	0.08	0.11
Other	0.04	0.03	0.05	0.06	0.04	0.08
White	0.45	0.39	0.51	0.74	0.71.	0.77
Lived with mother	0.85	0.82	0.88	0.88	0.87	0.90
Lived with father	0.48	0.45	0.50	0.62	0.59	0.65
Immigrant generation						
1st	0.09	0.07	0.11	0.06	0.04	0.07
2nd	0.13	0.11	0.16	0.11	0.10	0.13
3rd or greater	0.78	0.73	0.82	0.83	0.80	0.85
Household income	52 916	49 450	56 626	82 742	78 585	87 118
Maternal education						
Less than high school	0.14	0.12	0.17	0.05	0.04	0.06
High school graduate	0.51	0.48	0.55	0.42	0.40	0.44
Some college	0.22	0.20	0.25	0.26	0.24	0.28

Continued on next page

Table 2.1 – continued from previous page

	Disadvantaged (n=3,597)			Non-disadvantaged (n=6,003)		
	Mean	95% CI		Mean	95% CI	
College graduate	0.12	0.10	0.14	0.26	0.24	0.29
Maternal age at birth	24.58	24.13	25.03	26.83	26.52	27.14
Region						
Northeast	0.16	0.05	0.26	0.19	0.15	0.22
Midwest	0.13	0.06	0.20	0.26	0.22	0.30
South	0.51	0.42	0.60	0.3	0.24	0.37
West	0.2	0.15	0.25	0.25	0.19	0.31
Urbanicity						
Urban center	0.38	0.29	0.47	0.49	0.43	0.55
Urban fringe	0.41	0.31	0.51	0.38	0.31	0.45
Non-urban	0.21	0.11	0.30	0.13	0.10	0.16
Emotional disorder	0.28	0.25	0.32	0.24	0.22	0.25

We examined the associations between disadvantaged neighborhood residence and odds of emotional disorder by level of urbanicity by running unadjusted and adjusted models, both incorporating the survey design and weights (shown in Table 2.2). The adjusted model also uses the propensity score methods described above to control for potential confounding.

Table 2.2: Relative odds of emotional disorder comparing residence in disadvantaged versus non-disadvantaged neighborhoods by urbanicity.

	Unadjusted, weighted[1]		Propensity score subclassification, weighted	
	OR	95% CI	OR	95% CI
Non-urban	1.102	0.784, 1.549	0.707	0.467, 1.070
Urban fringe	1.103	0.816, 1.492	1.005	0.795, 1.270
Urban center	1.697	1.233, 2.336	1.591[b]	1.250, 2.025

[a] Based on all subjects (N=10 074). Results did not change when based on the 9600 subjects included in the primary analysis.

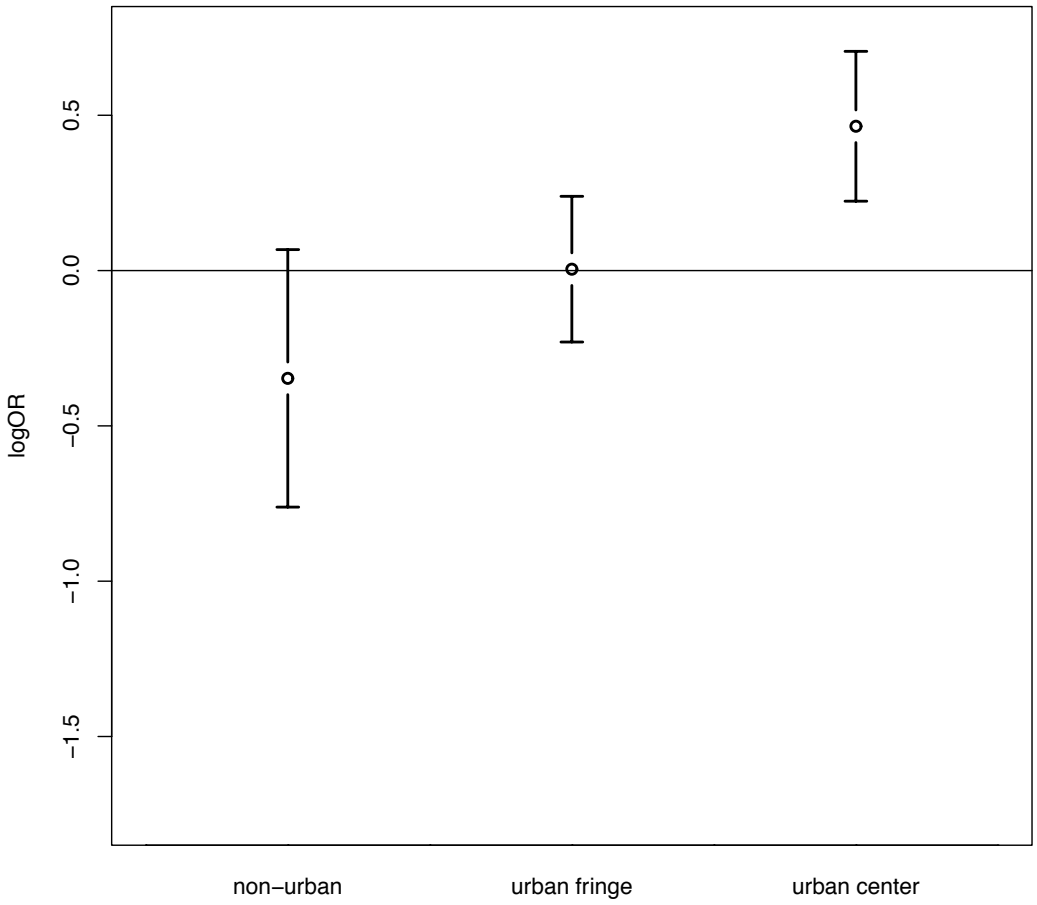
[b] The interaction term comparing the odds ratios between urban centers and non-urban areas was statistically significant (ratio of odds ratios: 2.08, 95% CI: 1.23, 3.55).

In the unadjusted model, living in a disadvantaged versus non-disadvantaged neighborhood within an urban center was associated with a 70% increased odds of emotional disorder (OR: 1.70, 95% CI: 1.23, 2.34). The unadjusted association between emotional disorder and neighborhood disadvantage was not statistically significant in urban fringe or non-urban areas.

The inferences remained the same in the adjusted model. Living in a disadvantaged versus non-disadvantaged neighborhood was associated with a 59% increased odds of emotional disorder (OR: 1.59, 95% CI: 1.25, 2.03) among those living within an urban center. The adjusted association between emotional disorder and neighborhood disadvantage was not statistically significant in urban fringe or non-urban areas.

Figure 2.3 depicts each odds ratio and its associated 95% confidence interval. There is a dose-response relationship between neighborhood disadvantage and odds of emotional disorder with higher odds ratios across increasing levels of urbanicity. The formal statistical test for interaction of the urban center and neighborhood disadvantage terms was statistically significant; the association between neighborhood disadvantage and emotional disorder is more than twice as large (ratio of odds ratios: 2.08, 95% CI: 1.23, 3.55) for adolescents living in urban centers versus non-urban areas. However, there is no statistically significant difference in the association between urban fringe and non-urban areas.

Figure 2.3: Log odds ratios and 95% confidence intervals for the effect of neighborhood disadvantage within strata of urbanicity for emotional disorders.



2.5 Discussion

In a large, nationally representative sample of U.S. adolescents, we found that urbanicity modified the association between neighborhood disadvantage and emotional disorder. Disadvantaged neighborhood residence was associated with emotional disorder if the neighborhood was within an urban center, but there was no association if the neighborhood was within a rural or urban fringe area. These results advance

previous research that did not consider the potentially modifying effect of urbanicity, and in part, address their conflicting results through demonstrating the impact of the urban environment. [12, 19, 23, 35, 42]

We recognize that the measurement of neighborhood disadvantage may differ depending on whether the neighborhood is in an urban or rural area. This issue is typically called measurement variance, and we performed a sensitivity analysis to assess whether our results could be an artifact of measurement variance of neighborhood disadvantage across levels of urbanicity. Multiple-group confirmatory factor analysis that allowed the loading coefficients of the neighborhood disadvantage measurement model to differ by urbanicity was used to estimate factor scores using the regression method. [2] Defining neighborhood disadvantage based on these factor scores did not change our inferences (results not shown but available from the first author).

As discussed in the introduction, urbanicity may exacerbate the association between neighborhood disadvantage and emotional disorders, because risk factors of poor mental health may be more prevalent in urban disadvantaged neighborhoods than in non-urban disadvantaged neighborhoods. A meta-analysis of 20 population-based surveys of adults in developed countries found 21% greater odds of anxiety disorders in urban areas compared to rural areas and 38% greater odds of mood disorders (e.g., major depressive disorder, bipolar disorder). [28] Despite a long history of research into the association between urbanicity and mental health, there has been surprisingly little research regarding specific characteristics of urban living that confer increased risk of emotional disorders. [9] Possibilities include lack of green space, residential instability, noise, and exposure to violence [e.g., [9, 21, 36]].

Our research findings have implications for future work and policy. Previous research suggested that children and adolescents in disadvantaged neighborhoods had

poorer mental health than those in more advantaged neighborhoods [e.g., [19, 23, 42]]. Our findings add that adolescents in disadvantaged urban areas may be particularly vulnerable to anxiety and depression, and therefore, targeting resources to this subpopulation may be appropriate. Risk conferred by the neighborhood and urban environment argues for policies that aim to improve neighborhood conditions. [26] Future work should identify specific neighborhood conditions that are detrimental to or protective of adolescent mental health. These findings can then be used to inform specific policy interventions such as crime prevention strategies, school climate and safety interventions, and housing policies such as those designed to reduce racial and economic segregation.

In addition, future research should examine whether urbanicity modifies the associations between neighborhood conditions and adolescent behavioral disorders and substance misuse. Although researchers have studied the relationship between neighborhood conditions and substance use, externalizing problems, and risky sexual behavior [e.g., [7, 13]], few have used behavioral disorders that correspond to DSM criteria. In addition, there has been little research in exploring possible effect heterogeneity.

Our analysis is subject to several assumptions and limitations. We assume that the sampling weights correctly account for sample selection and non-response, and thus the results generalize to the population of U.S. adolescents. We also assume that the propensity score model is correctly specified with no omitted variables.

2.5.1 Sensitivity analysis for an unobserved confounder

Omitted variables could result in biased effect estimates if there is confounding beyond that for which we have controlled. We used two bias equations for the odds ratio proposed by VanderWeele and Arah to assess the sensitivity of our results to an unobserved confounder. [39] For each equation, we made the following three assumptions as discussed by VanderWeele and Arah: (1) the association between the outcome and unobserved confounder is consistent across levels of exposure and observed covariates; (2) the unobserved confounder is binary; and (3) the prevalence of the unobserved confounder is constant across levels of the covariates. Let A denote the exposure received by an adolescent. In this study, $a1$ represents residence in a disadvantaged neighborhood and $a0$ represents residence in a non-disadvantaged neighborhood. Let Y denote the observed outcome, past 12-month emotional disorder present or absent. Let X denote the observed covariates, and let U denote the unobserved binary confounder.

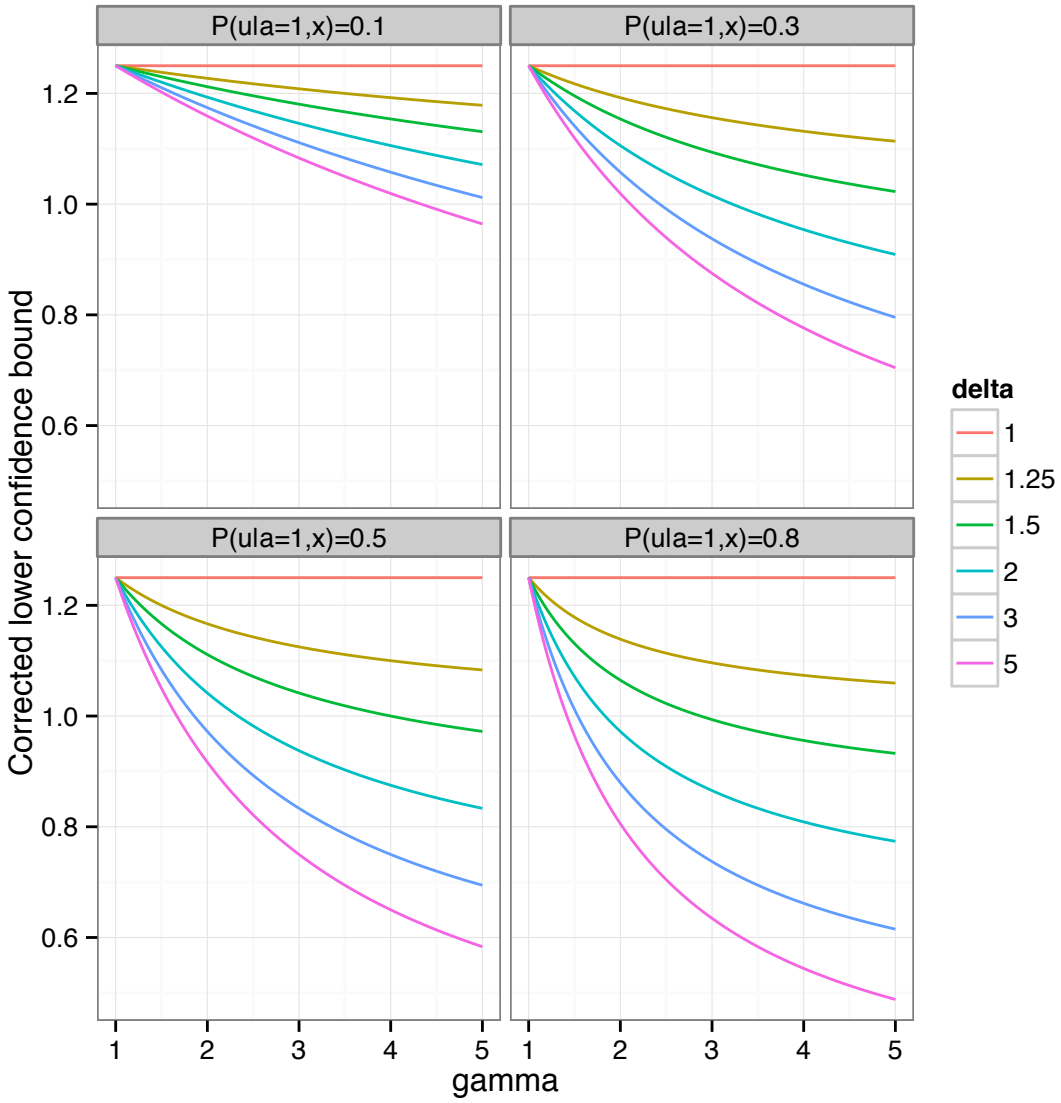
First, we used VanderWeele and Arah′s bias equation for the conditional odds ratio that makes the simplifying rare disease assumption.

$$d^{OR}(x) \approx \frac{1 + (\gamma - 1)P(u = 1|a = 1, x)}{1 + (\gamma - 1)P(u = 1|a = 0, x)} \tag{2.1}$$

We calculated the bias and resultant corrected lower 95% confidence bound across an array of input parameter values. We allowed the association between the outcome and unobserved confounder, $E(Y|u = 1, a)/E(Y|u = 0, a) = \gamma$, to range from 1 to 5. And the association between the exposure and the unobserved confounder $P(u = 1|a = 1, x)/P(u = 1|a = 0, x) = \delta$ also ranged from 1 to 5. The prevalence of the unobserved confounder in the exposed population ranged from 0.1 to 0.8.

Figure 2.4, below, plots the corrected lower 95% CI bound (on the y-axis) against the value of gamma (on the x-axis). The different colored curves in each subplot show the relationship for different values of delta. Each of the four subplots corresponds to different input values for the prevalence of the unobserved confounder in the exposed population (e.g., 0.1, 0.3, 0.5, and 0.8). For δ=2 and $P(u|a = 1, x)$=0.3, γ would have to be at least 3.3 to render our effect estimate nonsignificant.

Figure 2.4: Approximations of the corrected lower 95% confidence bound by values of delta, gamma, and $\mathrm{P}(u|a=1,x)$ making the rare disease assumption.



Second, we used VanderWeele and Arah′s bias equation for the conditional odds

ratio without making the simplifying rare disease assumption.

$$d^{OR}(x) = \left(\frac{\frac{E(Y|a=1,x,u=1)P(u=1|a=1,x)+E(Y|a=1,x,u=0)P(u=0|a=1,x)}{(1-E(Y|a=1,x,u=1))P(u=1|a=1,x)+(1-E(Y|a=1,x,u=0))P(u=0|a=1,x)}}{\frac{E(Y|a=0,x,u=1)P(u=1|a=0,x)+E(Y|a=0,x,u=0)P(u=0|a=0,x)}{(1-E(Y|a=0,x,u=1))P(u=1|a=0,x)+(1-E(Y|a=0,x,u=0))P(u=0|a=0,x)}} \right) \Big/ \left(\frac{\frac{E(Y|a=1,x,u=1)P(u=1|x)+E(Y|a=1,x,u=0)P(u=0|x)}{(1-E(Y|a=1,x,u=1))P(u=1|x)+(1-E(Y|a=1,x,u=0))P(u=0|x)}}{\frac{E(Y|a=0,x,u=1)P(u=1|x)+E(Y|a=0,x,u=0)P(u=0|x)}{(1-E(Y|a=0,x,u=1))P(u=1|x)+(1-E(Y|a=0,x,u=0))P(u=0|x)}} \right) \tag{2.2}$$

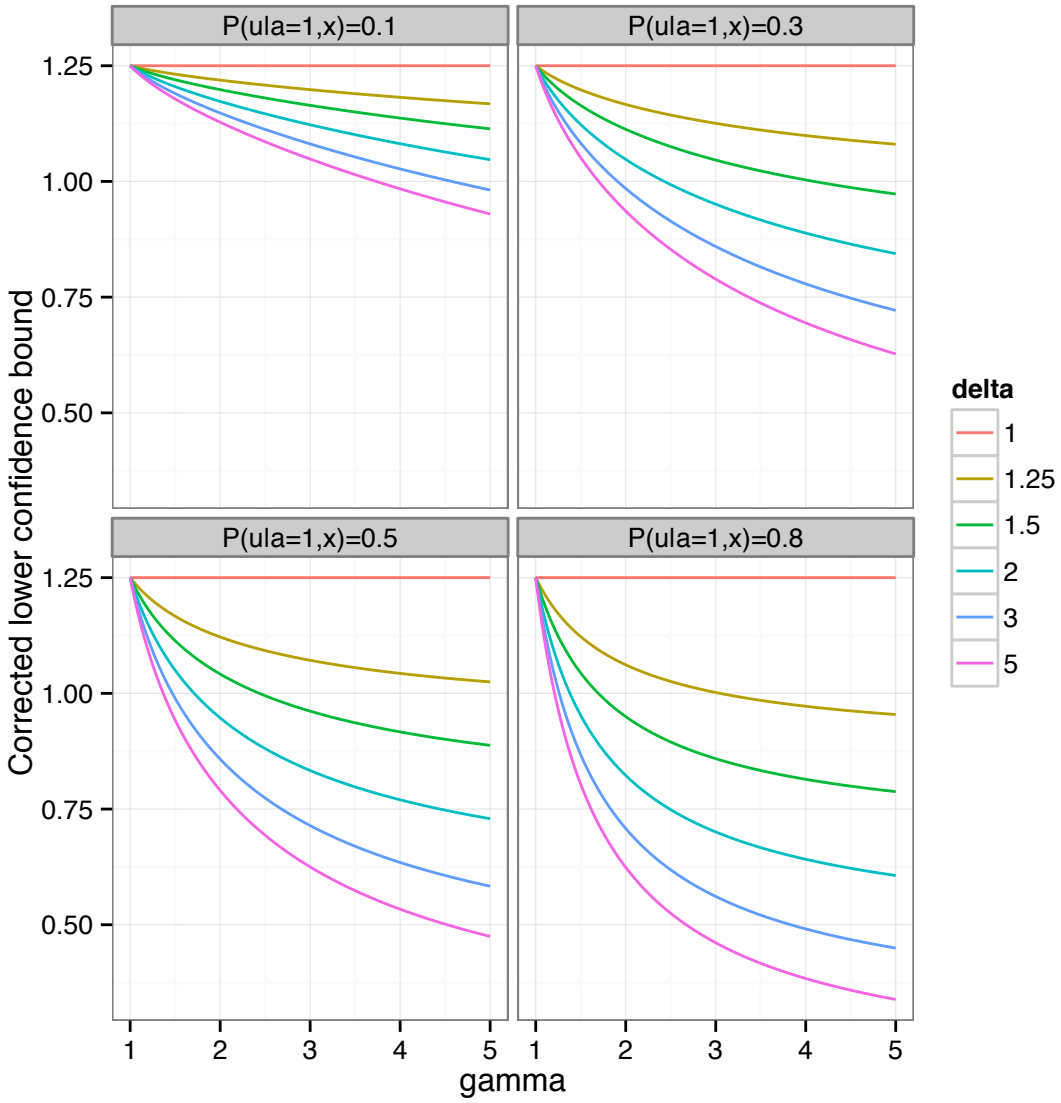
The rare disease assumption is considered appropriate if the prevalence is less than ten percent. The unadjusted prevalence of emotional disorder in the population of adolescents living in non-disadvantaged neighborhoods was 0.24.

We specified input values for γ, δ, and $P(u = 1|a = 1, x)$ as detailed above. However, the exact calculation of the conditional odds ratio bias necessitated the specification of two additional parameters. First, we set $E(Y = 1|u = 1, a = 1, x) = 0.5$. Second, as defined by our dichotomization of the exposure, we set the prevalence of exposure to one-third. This was used to calculate $P(u|x)$.

$$P(u|x) = P(u|a = 1, x)P(a = 1|x) + P(u|a = 0, x)P(a = 0|x) \tag{2.3}$$

Similar to Figure 2.4, Figure 2.5, below, plots the corrected lower 95% confidence bound against the value of γ. The different colored curves in each subplot show the relationship for different values of delta. Each of the four subplots corresponds to different input values for the prevalence of the unobserved confounder in the exposed population (e.g., 0.1, 0.3, 0.5, and 0.8).

Figure 2.5: Estimates of the corrected lower 95% confidence bound by values of delta, gamma, and $\mathrm{P}(u|a=1,x)$ using the exact equation.



As seen in both figures, our results are more sensitive to a prevalent unobserved confounder. In the exact analysis, when $P(u|a=1,x)$=.8, a U with a 2-fold association with A and a 1.4-fold association with Y would change our inference. When $P(u|a=1,x)$=.3, a U with a 2-fold association with A and a 2.5-fold association with Y would change our inference.

We also assume that measurement error is minimal and does not affect our inferences. However, our exposure variable is subject to measurement error for at least two reasons. First, Census tract is a proxy for neighborhoodâĂŤhow residents and/or city planners would map neighborhood boundaries may differ from Census tract boundaries. Nevertheless, Census tracts allow neighborhood measures to be "compared over time and across regions", [20] and are better than zip codes at detecting differences in socioeconomic gradation across areas. [20] Second, we may have misclassified neighborhoods as disadvantaged or non-disadvantaged. However, we believe that this misclassification is likely minimal as we are using a previously established scale and dichotomizing the exposure. Although using the first tertile as the cut-point may not be optimal, misclassification would likely be non-differential and bias our estimates toward the null.

The outcome variable of emotional disorder is also subject to measurement error. However, the use of the CIDI in assessing mental disorder is a key strength. It is more reliable than unstandardized psychiatric diagnoses and shows good agreement with standardized psychiatric diagnoses. [41] It also has high content validity, as it is designed to correspond to DSM-IV and ICD-10 criteria. [16]

The large, nationally representative sample is a major strength of this analysis. With over 10,000 adolescents, the NCS-A is the largest nationally representative survey of adolescent mental health in the U.S., compiling data collected from adolescents, parents, and GIS-coded residence for an unusually large amount of information on context. Because of these attributes, we were well positioned to examine the role of urbanicity as a potential effect modifier of the association between neighborhood disadvantage and mental health. Incorporation of the survey design and weights into our analysis preserves the NCS-A's sampling strengths. It accounts for sample selection

(incorporating strata and cluster variables) and nonresponse, thus addressing clustering of adolescents within neighborhoods. (Clustering was low in this sample—there were an average of three adolescents per neighborhood). Incorporation of the survey design and weights allows us to interpret the results as being nationally representative, reduces bias from differential non-response, and helps protect our inferences from inflated Type-1 error rates through better standard error estimation. [10]

Another strength is our use of propensity score subclassification. As discussed previously, propensity score methods have an advantage over standard regression methods in that they allow the analyst to look at the data to assess (1) how well bias is controlled for in each covariate and (2) the extent of propensity score overlap and thus the range at which the data will support estimates. However, analysis using a standard multivariate logistic regression model did not change our inferences (results not shown but available from the first author). In addition, we could have used other methods besides propensity score subclassification such as inverse probability of treatment weighting (IPTW). However, IPTW has well-established efficiency concerns, [29] and multiplying inverse probability of treatment weights by the large survey weights may further exacerbate these issues. Further, there is some empirical evidence that subclassifcation may perform slightly better than weighting when combined with complex survey data. [6]

2.5.2 Conclusion

The associations between urbanicity and mental health and between neighborhood disadvantage and mental health have been studied separately for more than 100 years. [5, 8] No known research has studied effect modification of the associa-

tion between neighborhood disadvantage and mental health by urbanicity. However, adolescents live in neighborhoods that simultaneously comprise a level of urbanicity and a level of disadvantage. Our results suggest that the effect of neighborhood disadvantage on adolescent depression and anxiety is greater in urban centers than in non-urban areas. Recognizing the dependence of these contexts will aid future research in identifying specific characteristics of urban, disadvantaged neighborhoods (e.g., violence, residential instability, lack of green space) that confer risk of mental disorder, and in this era of shrinking budgets, will help channel funding and services to those youth most at risk.

Chapter 3

The association between cortisol and neighborhood disadvantage in a U.S. population-based sample of adolescents

This chapter's research has been published. [64] All code associated with this chapter can be found here:
https://github.com/cherrygarcia/Aim2.

3.1 Abstract

The association between neighborhood conditions and cortisol is rarely studied in children or adolescents and has been hampered by small sample size and racial/ethnic and geographic homogeneity. Our objective was to estimate the association between neighborhood disadvantage and salivary cortisol levels in a large, geographically and racially/ethnically diverse sample of adolescents from the National Comorbidity Survey Replication Adolescent Supplement. Salivary cortisol was collected before and after an interview administered in the adolescent's home. We used a propensity score approach to match adolescents living in disadvantaged neighborhoods with those in non-disadvantaged neighborhoods to create two similar groups based on the time and

day of cortisol collection as well as demographic characteristics. Adolescents living in disadvantaged neighborhoods had higher pre-interview cortisol levels and steeper rates of decline in cortisol levels over the course of the interview than similar adolescents in non-disadvantaged neighborhoods. This bolsters the evidence base suggesting that place may influence the stress response system.

3.2 Introduction

Place may influence health through several pathways; stress is one potential mediator that is frequently invoked. [3, 18] For example, living in a blighted urban neighborhood may increase exposure to stressors such as violence, noise, and crowding. These exposures may elicit repeated activations of the stress response system, which in turn may lead to eventual dysregulation of the hypothalamic-pituitary-adrenal (HPA) axis, a primary stress regulatory system. HPA axis dysregulation has been associated with a range of mental, cardiovascular, immunologic, and metabolic disorders. [3,66] Although recent studies have found associations between neighborhood conditions and stress biomarkers, [5,11,19,54] there has been limited research on links between neighborhood conditions and stress biomarkers in children or adolescents.

Cortisol is a hormone involved in the HPA axis [51] that has been used in several contexts. Adverse conditions in neighborhood and family environments have been linked to both cortisol levels and cortisol reactivity, although the evidence is mixed. In adults, some studies have yielded associations between neighborhood- and individual-level low socioeconomic status (SES) and cortisol diurnal levels—specifically lower waking levels, [29] higher average levels, [13,14] and less steep declines over the course of the day, [1,19,29,36] though others have found null or opposite results. [1,13,29] Do

et al. also found that neighborhood violence was associated with lower cortisol levels at awakening and less steep initial declines. [19] In children, studies have reported associations between individual-level disadvantage (including low SES, exposure to stressful life events, and family adversity) and lower morning cortisol levels, [4, 58] higher average cortisol levels, [4, 21, 37, 58] and less steep declines. [26, 37, 50] In addition, some have suggested a curvilinear (upside-down u-shaped) association; children and adolescents exposed to the most stressful conditions have cortisol levels that resemble those of non-disadvantaged individuals. [4, 26]

The link between individual- and neighborhood-level adverse conditions and cortisol reactivity is likely complex. Some studies have shown that adverse conditions in childhood are associated with greater cortisol reactivity in adulthood, [24, 49, 56] but lifetime adversity is associated with blunted reactivity. [24, 46] Others have found no association. [67] Relatedly, moderate adversity has been associated with heighted reactivity in children and adolescents, [27, 58] whereas more severe forms of adversity, such as prolonged child maltreatment, has been associated with blunted reactivity. [48] The timing and duration of exposure to adverse conditions may also be influential. [7, 67]

The evidence for an independent association between adverse neighborhood conditions and salivary cortisol in adolescents is extremely limited. Studies conducted to date provide preliminary evidence that neighborhood disadvantage is associated with higher average resting cortisol levels [9, 11] and greater cortisol reactivity. [28] However, the studies have been based on small, racially homogeneous samples in single urban areas. [9, 11, 28] The present study was motivated to address this gap in the literature. We used the National Comorbidity Survey Replication Adolescent Supplement (NCS-A) to estimate the association between neighborhood disadvan-

tage and salivary cortisol levels in adolescents. The NCS-A consists of a nationally representative, ethnically diverse sample of adolescents in the United States. Cortisol measurements are available for 2490 of the adolescents, making it the largest sample of cortisol in U.S. children or adolescents. Our analyses of these data utilize a propensity score approach coupled with regression adjustment designed to address a key threat to internal validity—non-random neighborhood assignment and consequent imbalance of confounding variables, including those particularly influential to cortisol measurement.

3.3 Methods

3.3.1 Study sample

The NCS-A is a nationally representative, cross-sectional survey of U.S. adolescent mental health. The background, design, sampling, and field procedures are presented elsewhere. [38–40,53] Participants ages 13-17 were recruited from a dual-frame sample consisting of household and school subsamples. Face-to-face, computer-assisted interviews (which included a modified Composite International Diagnostic Interview) were conducted in the adolescent's home by professional interviewers from the Survey Research Center at the Institute for Social Research at the University of Michigan. The interviews took place between February 2001 and January 2004. While the adolescent was interviewed, his/her parent or parent surrogate was given a self-administered questionnaire. A short-form version of the questionnaire was administered to parents who did not complete the long-form version. Each participating adolescent and his/her parent or guardian provided informed assent and consent. Recruitment and

consent procedures were approved by the Human Subjects Committees of Harvard Medical School and the University of Michigan.

3.3.2 Contextual measures

3.3.2.1 Neighborhood Disadvantage

The Survey Research Center at the Institute for Social Research at the University of Michigan geocoded residential addresses to U.S. Census tracts. Neighborhood SES, defined at the Census tract level, is a summary measure created by [17] that has been used previously. [6, 32, 55, 61] We defined neighborhoods in the lowest SES tertile as disadvantaged, and those in the middle and upper tertiles as non-disadvantaged.

Neighborhood SES is made up of six indicators from the U.S. Census Short Form 3 (SF3): log median household income; % households with interest, dividend, or rental income; log median value of housing units; % persons over age 25 with high school degree; % persons over age 25 with college degree; % persons in executive, managerial, or professional specialty occupations. [17] The normally distributed summary score results from summing the z-score of each indicator. In the NCS-A sample, the summary score has a median value of -0.36 (range: -13.6, 17.8) and a Cronbach's alpha of 0.83.

3.3.3 Individual Measures.

3.3.3.1 Outcome measures

Cortisol levels in ng/mL were measured using saliva samples. Saliva samples were collected in a salivette by passive drool after the participant chewed on a piece of sugarless gum immediately before and after the interview, while the interviewer was present. The interviewer's laptop automatically recorded the time and date of each sample collection, and interviews lasted an average of 149 minutes. Salivettes were treated with sodium azide at Harvard University, centrifuged, and pre-labeled with subject identification numbers and study information prior to sample collection. After collection, samples were mailed to NIH where they were stored at -80C until testing. Quantification of cortisol levels was done by a radioimmunoassay (Siemens Diagnostic). The sensitivity of the assay was 0.0165 ng/mL. Intra- and inter-assay coefficients of variation were 5.4% and 26.0%, respectively. Similar coefficients of variation for this method have been reported previously. [71]

We examined three outcomes in the present study: (1) point-in-time pre-interview cortisol level, (2) point-in-time post-interview cortisol level, and (3) cortisol rate of change (slope) over the course of the interview, calculated as the difference in post versus pre-interview levels divided by the length of the interview in hours. These outcomes do not directly map onto specific HPA axis dimensions. For example, a Trier stress test [42] measures stress reactivity, but such a test was deemed inappropriate for children by the NCS-A investigators. Instead, pre- and post-interview samples measure cortisol in slightly different naturalistic settings. In the case of the pre-interview sample, the adolescent is interrupted from his/her normal routine for the interview (the adolescent may have been active—not sitting quietly), and the

adolescent is anticipating the new experience of being interviewed by a stranger as part of a survey of mental health (which may provoke an HPA response in some). In the case of the post-interview sample, the adolescent has been seated for an average of $2\ \frac{1}{2}$ hours and has finished answering questions that were designed not to be stressful. We hypothesized that the extent to which neighborhood disadvantage is associated with cortisol may differ slightly in these two settings. In addition, we hypothesized that cortisol levels would decline over the course of the interview due to cortisol's circadian rhythm, the seated nature of the interview, and the fact that the questions were designed not to be stressful, and that the rate of decline may also be associated with neighborhood disadvantage.

Cortisol was measured on the first 2490 adolescents enrolled in the NCS-A. Budget limitations prevented cortisol testing on the full sample. Participants with and without cortisol measures were similar across most demographic characteristics (see Table 3.1). They differed slightly in terms of where they were sampled (e.g., region of the country, urbanicity, neighborhood disadvantage status), when they were sampled (e.g., season, time of interview), and in terms of family income, maternal education, current parental employment, small for gestational age, and smoking status. Five of the 2490 adolescents with cortisol data were excluded, because they resided in Census tracts with missing or inestimable SF3 indicators, which precluded characterization of their neighborhood. Removing extreme outlying cortisol values resulted in the exclusion of six additional participants. A small proportion of adolescents were not able to complete the interview in one sitting. In an effort to reduce heterogeneity in the post-interview measure, we excluded those adolescents who took more than four hours to complete the interview (2.4%) from the post-interview and rate of change outcome analyses. However, inclusion of those adolescents did not change the infer-

ences (results not shown).

Table 3.1: NCS-A sample characteristics in 2001-2004 by cortisol status. Results are combined across imputations and survey design-based standard errors are estimated using Taylor linearization.

Variable	**Cortisol N=2485**		**No cortisol N=7589**		**P-value (2-sided)**
	Mean	**SE**	**Mean**	**SE**	
Female, %	49.46	0.97	51.61	0.60	0.065
Age, y	15.153	0.036	15.186	0.032	0.248
Race/ethnicity, %					0.710
Hispanic	18.87	1.16	18.75	0.83	
Black	18.71	1.25	19.48	0.95	
Other	6.52	0.57	6.01	0.34	
White	55.90	1.54	55.76	1.173	
Urbanicity, %					0.001
Non-urban	23.18	2.20	22.73	1.71	
Suburb	35.33	1.89	31.93	1.46	
Urban center	41.49	1.88	45.34	1.53	
Region, %					<0.001
Northeast	16.42	1.43	18.94	1.20	
Midwest	25.15	1.80	28.32	1.50	
South	36.78	2.04	33.05	1.53	

Continued on next page

Table 3.1 – continued from previous page

Variable	During CAR		Post-CAR		P-value
West	21.65	1.48	19.70	1.14	
Household income (log)	11.106	0.026	11.188	0.0167	0.005
Maternal age at birth, y	26.034	0.128	26.049	0.086	0.463
Maternal education, %					0.012
Less than high school	10.78	1.10	8.79	0.45	
High school graduate	43.53	1.12	43.66	0.69	
Some college	24.00	1.02	23.95	0.60	
College graduate	21.69	1.06	23.59	0.64	
Maternal work history, %					0.461
All of adolescent's life	48.31	1.05	49.53	0.66	
Most of adolescent's life	21.94	0.84	20.73	0.46	
Some of adolescent's life	14.98	0.73	14.15	0.43	
A little of adolescent's life	7.09	0.54	7.41	0.32	
Not at all	7.68	0.57	8.17	0.35	
Paternal work history, %					0.867
All of adolescent's life	77.29	1.05	76.72	0.64	
Most of adolescent's life	13.03	0.83	13.71	0.44	
Some of adolescent's life	6.20	0.62	5.99	0.33	
A little of adolescent's life	2.16	0.37	2.08	0.24	
Not at all	1.32	0.29	1.49	0.19	
Parent current employed, %	72.91	4.57	75.08	0.81	0.033

Continued on next page

Table 3.1 – continued from previous page

Variable	During CAR		Post-CAR		P-value
Family structure, %					
Lived with mother whole life	87.70	0.68	86.83	0.42	0.276
Lived with father whole life	57.20	1.12	56.16	0.74	0.378
Citizen, %	95.90	0.44	95.63	0.27	0.604
English as second language, %	20.60	1.09	19.86	0.77	0.437
Immigrant generation, %					0.519
1st	5.98	0.56	5.88	0.34	
2nd	12.27	0.80	13.16	0.55	
3rd or greater	81.74	1.02	80.96	0.73	
Small for gestational age, %	6.45	0.67	5.33	0.42	0.040
Current smoker, %	5.72	0.48	7.29	0.35	0.009
Current drug user, %	5.51	0.47	6.01	0.31	0.389
Current oral contraceptive, %	3.66	0.37	4.59	0.27	0.057
No. Rx	0.4348	0.020	0.4559	0.012	0.188
Typical bedtime,hr					
Weekday	00:22	00:02	00:19	00:02	0.617
Weekend	22:26	00:02	22:25	00:01	0.850
Typical hours of sleep,hr					
Weekday	7.672	0.030	7.676	0.018	0.460
Weekend	8.845	0.044	8.851	0.027	0.454
Physical abuse by parent, %	12.30	0.69	12.74	0.45	0.594

Continued on next page

Table 3.1 – continued from previous page

Variable	During CAR		Post-CAR		P-value
3+ parental adversities, %	3.70	0.39	3.83	0.24	0.811
Disadvantaged neighborhood, %	38.07	2.02	35.37	1.53	0.016
Cortisol sample measurements					
Season, %					<0.001
Spring	10.54	0.89	27.54	0.89	
Summer	42.05	1.32	27.99	0.80	
Fall	33.08	1.24	24.42	0.87	
Winter	14.33	0.87	20.06	0.71	
Weekend, %	29.18	0.97	29.62	0.64	0.690
Collection time, hr	14:52	00:04	15:02	00:02	<0.001

Cortisol peaks in the morning during the cortisol awakening response (CAR) and then declines over the rest of the day. [12] Research suggests that the association between salivary cortisol levels and SES may be in opposite directions depending on whether cortisol is measured during the CAR or afterward. [19] To avoid this source of heterogeneity, we limited our analysis to those adolescents with measures taken during the late decline portion of the cortisol diurnal cycle. [23] We operationalized this by excluding adolescents whose first sample was taken prior to 10am on a weekday during the school year and prior to noon on a weekend or during summer vacation (we did not have information on the time the adolescent awoke on the day of the interview). Participants with cortisol measures during versus after the CAR were

similar across most demographic characteristics (see Table 3.2). There were slight differences in season, weekend versus weekday, typical weekend bedtime, and family structure.

Table 3.2: NCS-A cortisol sample characteristics by CAR sampling time. Results are combined across imputations and survey design-based standard errors are estimated using Taylor linearization.

Variable	**During CAR N=449**		**Post-CAR N=2036**		**P-value (2-sided)**
	Mean	**SE**	**Mean**	**SE**	
Female, %	47.66	2.33	49.85	1.07	0.430
Age, y	15.116	0.073	15.161	0.038	0.290
Race/ethnicity, %					0.497
Hispanic	18.49	2.07	18.96	1.19	
Black	20.71	2.17	18.27	1.31	
Other	7.35	1.23	6.34	0.62	
White	53.45	2.61	56.43	1.61	
Urbanicity, %					0.063
Urban center	22.72	2.88	23.28	2.24	
Suburb	31.18	2.70	36.25	1.98	
Non-urban	46.10	2.92	40.47	1.92	
Region, %					0.705

Continued on next page

Table 3.2 – continued from previous page

Variable	During CAR		Post-CAR		P-value
Northeast	16.48	2.16	16.40	1.47	
Midwest	26.73	2.55	24.80	1.87	
South	34.52	2.96	37.28	2.10	
West	22.27	2.30	21.51	1.53	
Household income (log), dollars	11.135	0.052	11.100	0.028	0.724
Maternal age at birth,y	26.109	0.281	26.018	0.140	0.614
Maternal level of education, %					0.575
Less than high school	9.29	1.87	11.11	1.22	
High school graduate	43.83	2.95	43.46	1.19	
Some college	25.81	2.33	23.61	1.13	
College graduate	21.07	2.11	21.83	1.17	
Maternal work history, %					0.729
All of adolescent's life	51.29	2.42	47.65	1.16	
Most of adolescent's life	20.96	2.08	22.16	0.93	
Some of adolescent's life	13.81	1.63	15.24	0.82	
A little of adolescent's life	6.84	1.21	7.14	0.59	
Not at all	7.10	1.22	7.81	0.65	
Paternal work history, %					0.942
All of adolescent's life	76.61	2.42	77.44	1.12	
Most of adolescent's life	13.07	1.77	13.03	0.89	
Some of adolescent's life	6.19	1.28	6.20	0.67	

Continued on next page

Table 3.2 – continued from previous page

Variable	During CAR		Post-CAR		P-value
A little of adolescent's life	2.67	0.96	2.04	0.40	
Not at all	1.45	0.72	1.30	0.29	
Parental current employment, %	69.24	5.58	73.71	4.50	0.061
Family structure, %					
Lived with mother whole life	87.53	1.59	87.74	0.75	0.964
Lived with father whole life	61.51	2.32	56.25	1.24	0.047
Citizen, %	94.88	1.07	96.12	0.50	0.285
English as second language, %	20.94	2.15	20.53	1.13	0.898
Immigrant generation, %					0.858
1st	6.50	1.27	5.87	0.62	
2nd	11.92	1.61	12.35	0.88	
3rd or greater	81.58	2.03	81.78	1.08	
Small for gestational age, %	6.50	1.63	6.44	0.67	1.000
Current smoker, %	4.23	0.95	6.05	0.54	0.164
Current drug user, %	4.01	0.92	5.84	0.54	0.153
Current oral contraceptive user, %	4.90	1.01	3.39	0.39	0.160
No. Rx	0.452	0.046	0.431	0.022	0.663
Typical bedtime,hr					
Weekday	22:42	00:04	22:44	00:02	0.427
Weekend	00:13	00:05	00:24	00:02	0.023
Typical hours of sleep, hr					

Continued on next page

Table 3.2 – continued from previous page

Variable	During CAR		Post-CAR		P-value
Weekday	7.770	0.071	7.650	0.032	0.939
Weekend	8.868	0.102	8.839	0.047	0.602
Physical abuse by parent, %	10.47	1.54	12.71	0.75	0.220
3+ parental adversities, %	2.00	0.66	4.08	0.45	0.049
Disadvantaged neighborhood, %	38.53	3.00	37.97	2.07	0.866
Cortisol sample measurements					
Season, %					<0.001
Spring	7.80	1.36	11.15	1.00	
Summer	57.46	2.47	38.65	1.40	
Fall	24.72	2.21	34.92	1.33	
Winter	10.02	1.46	15.28	0.94	
Weekend, %	56.12	2.47	23.23	0.98	<0.001
Collection time, hr	10:12	00:03	15:54	00:04	<0.001

3.3.3.2 Covariate measures

We included covariates that have been recommended [43] as potentially important covariates to control for in assessing relationships between neighborhood and mental health in addition to covariates hypothesized to confound the association between neighborhood and cortisol.

By definition, confounding variables are not affected by the exposure of inter-

est. In contrast, mediators (sometimes called explanatory variables) lie on the causal pathway between exposure and outcome. With this distinction in mind, we controlled for confounding through the following variables: adolescent age (in years), race/ethnicity, urbanicity, immigrant generation, citizenship, English as a second language, maternal age at birth (in years), maternal level of education, region, time of cortisol collection, weekend versus weekday, and season. Although maternal age at birth and maternal education may have been affected by prior neighborhood residence (which may be related to current residence), we decided to include these variables to control for two important measures of family socioeconomic status. [43] Household income (log-transformed) and parental employment may be thought of as confounders or mediators. We controlled for these variables in a separate model. Potentially mediating variables included: small for gestational age status, body-mass index (BMI), typical bedtime, and typical hours of sleep. [8, 57, 60]

Although some variables (e.g., adolescent's age) had no missing data, others had missing data due to non-response. The two variables with the most missing data were maternal education (19% missing) and current parental employment (30% missing). Following recommendations for addressing missing data, [68] we performed multiple imputation by chained equations, [10] creating 10 imputed datasets. This approach may be less biased than one that restricts analyses to those with complete data. [33, 68] The imputation model included variables used in the analysis as well as those predictive of non-response.

3.3.4 Exclusion criteria

In addition to exclusions described in the Outcome Measures section, we excluded participants who were currently pregnant (0.3%), diagnosed with type-1 diabetes (0.9%), taking psychiatric medication (4.8%), possibly taking steroid mediation for asthma (3.6%), taking oral contraceptives (3.7%), current smokers (5.7%), and current illicit drug users (8.4%). After applying the exclusion criteria, there were 1646 adolescents in the sample. Four adolescents with a pre-interview measure did not have a post-interview measure. All adolescents that had a post-interview cortisol measure also had a pre-interview measure.

3.3.5 Analytic Approach

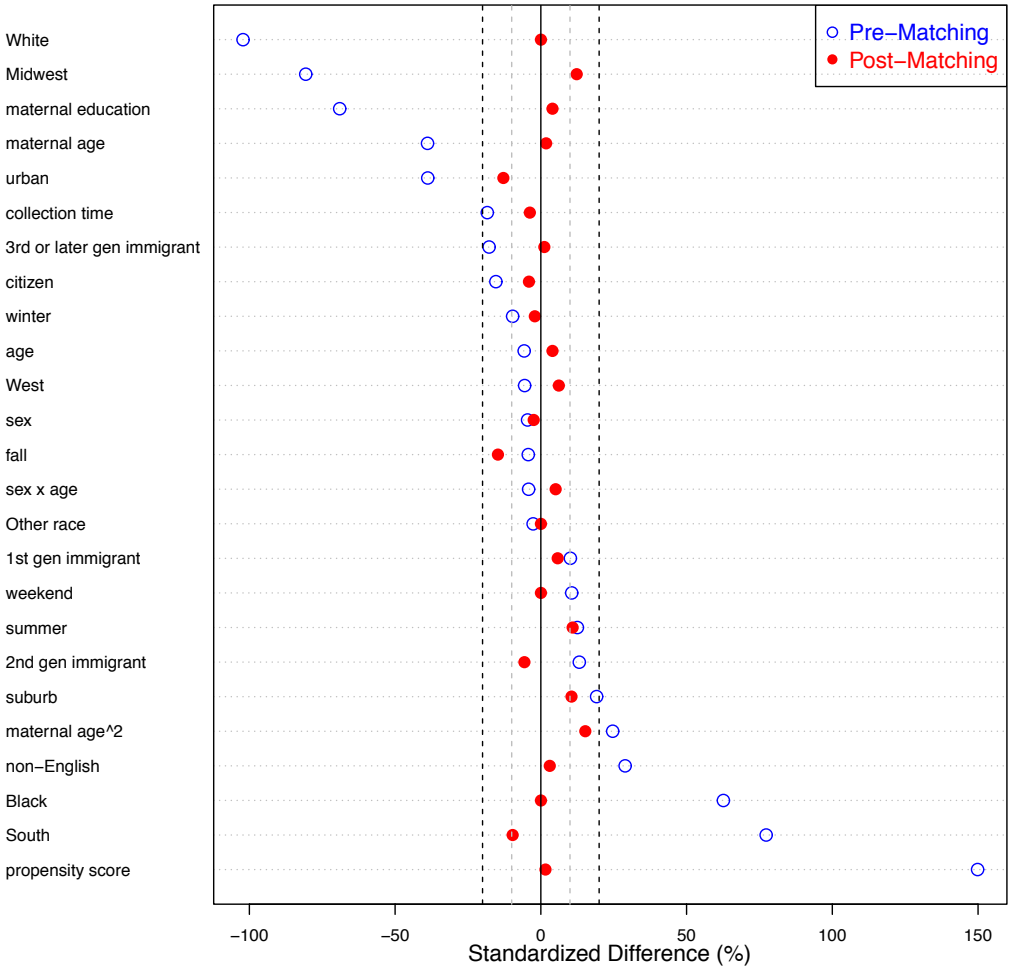
3.3.5.1 Matching

We used a propensity score approach to match adolescents living in disadvantaged neighborhoods with those in non-disadvantaged neighborhoods to create two similar groups based on the time and day of cortisol collection as well as demographic characteristics. We followed the recommendations of Rubin and Thomas and combined propensity score matching (with replacement) with the more stringent matching requirements of exact and caliper matching on a subset of particularly influential covariates. [63] Because research suggests that time of day, race, and weekend versus weekday are important determinants of cortisol levels and reactivity, [15,16,29,57] we exact matched on race/ethnicity and weekend/weekday and matched within calipers of time of day (0.20 standard deviations).

After exact and caliper matching, adolescents were matched based on their esti-

mated propensity scores. We estimated a propensity score [59] for each adolescent from a logistic regression model of living in a disadvantaged neighborhood at the time of the study as a function of the covariates detailed in Figure 3.1; the propensity scores are the resulting predicted probabilities from this model. Covariate balance between the disadvantaged versus non-disadvantaged neighborhood groups was achieved in the final matched dataset (all standardized mean differences less than 15%, Figure 3.1).

Figure 3.1: Covariate balance pre- and post-matching. Plotted points represent the standardized mean differences (difference in means between the disadvantaged neighborhood group and non-disadvantaged neighborhood group standardized by the standard deviation in the disadvantaged group) for each covariate. Open dots represent standardized mean differences in the pre-matched data. Closed dots represent standardized mean differences in the post-matched data. Vertical grey and black dashed lines indicate standardized mean differences of 10% and 20%, respectively. Participants were exact matched on race/ethnicity and weekend/weekday, caliper matched on time of sample, and matched on propensity score that was a function of the covariates listed on the y-axis of the figure.



3.3.5.2 Regression

After matching, between 852 and 894 adolescents remained in the sample, depending on imputation number. Table 3.3 describes the final matched sample by neighborhood disadvantage status. These adolescents resided in between 531 and 550 Census tracts, depending on imputation number. One way to address clustering of people in neighborhoods is through multilevel models; however, this strategy requires having enough adolescents within the same neighborhood to accurately estimate neighborhood parameters. Typical guidance suggests using multilevel models if there are at least 15-30 residents per neighborhood. [43] In this sample, clustering of participants within neighborhoods was low: there was an average of 1.6 adolescents per neighborhood, and only 10% of neighborhoods had more than two adolescent participants. Because multilevel modeling is not recommended in such scenarios, we accounted for clustering by using the survey package in R to incorporate sampling strata and cluster variables in standard error estimation using Taylor linearization). [47] This strategy produced similar results as calculating cluster robust standard errors using a sandwich estimator (results not shown but available from the first author upon request).

Table 3.3: NCS-A matched sample characteristics by neighborhood disadvantage status. Mean (SE)[2]

Variable	Disadvantaged neighborhood N=602	Non-disadvantaged neighborhood N=265	P-value (2-sided)
Female (/Age (y)	14.98 (0.10)	14.90 (0.15)	0.59
Race/ethnicity (%)			1
Hispanic	30.73 (3.37)	30.73 (6.32)	
Black	37.21 (3.49)	37.21 (6.54)	
Other	4.82 (1.13)	4.82 (1.82)	
White	27.24 (3.37)	27.24 (2.95)	
Urbanicity (%)			0.69
Urban center	30.40 (4.13)	35.22 (5.57)	
Suburb	42.02 (3.40)	38.37 (5.28)	
Non-urban	27.58 (3.00)	26.41 (7.60)	
Region (%)			0.55
Northeast	10.13 (2.65)	11.79 (2.08)	
Midwest	9.80 (1.61)	6.15 (1.85)	
South	60.80 (3.69)	64.29 (5.39)	
West	19.27 (2.80)	17.77 (3.48)	

Continued on next page

[2]Matched sample from the first imputation, matching weights used, clustering by neighborhood accounted for. P-values were calculated from the t statistic for continuous covariates and from the chi-squared statistic for categorical covariates.

Table 3.3 – continued from previous page

Variable	Disadvantaged	Non-disadvantaged	P-value
Household income (log, dollars)	10.40 (0.09)	10.63 (0.14)	0.17
Maternal age at birth (y)	24.82 (0.30)	24.77 (0.40)	0.92
Maternal level of education (%)			0.67
Less than high school	19.27 (1.59)	18.94 (5.03)	
High school graduate	49.67 (2.44)	54.65 (5.49)	
Some college	21.43 (1.48)	16.61 (2.22)	
College graduate	9.63 (1.42)	9.80 (2.05)	
Family structure (%)			
Lived with mother whole life	87.71 (1.07)	86.05 (3.67)	0.68
Lived with father whole life	47.34 (1.93)	46.35 (4.23)	0.83
Immigrant generation (%)			0.87
1st	8.64 (1.19)	9.63 (2.99)	
2nd	16.11 (1.51)	17.11 (3.35)	
3rd or greater	75.25 (2.17)	73.36 (4.02)	
Small for gestational age[3] (%)	8.80 (0.79)	6.64 (1.73)	0.31
Physical abuse by parent (%)	15.45 (1.54)	17.61 (4.06)	0.59
3+ parental adversities (%)	3.16 (0.61)	1.33 (0.69)	0.11
Season (%)			0.78
Spring	11.13 (2.22)	9.63 (2.60)	
Summer	42.19 (4.00)	39.87 (4.20)	

Continued on next page

[3]Defined using guidelines in [2]

Table 3.3 – continued from previous page

Variable	Disadvantaged	Non-disadvantaged	P-value
Fall	33.39 (3.15)	38.21 (4.52)	
Winter	13.29 (1.94)	12.29 (3.09)	
Weekend (%)	27.24 (2.43)	27.24 (4.72)	1
Interview time (hr)	15:37 (0:08)	15:38 (0:14)	0.97
Pre-interview cortisol (ng/mL×10-2)	25.20 (0.82)	20.68 (1.30)	<0.01
Post-interview cortisol (ng/mL×10-2)	14.07 (0.49)	16.21 (1.47)	0.15
Cortisol slope (ng/mL/hr×10-2)	-4.69 (0.33)	-1.84 (0.72)	<0.01

We ran fully parameterized outcome regression models using the final matched dataset and weighting by the matching frequency weights (to account for matching with replacement) to estimate associations between living in a disadvantaged versus non-disadvantaged neighborhood and (1) pre- and (2) post-interview cortisol levels and (3) cortisol slope, conditional on potential confounders. Although researchers may address confounding by either (1) including confounders in the propensity score matching model or (2) by including confounders in the regression model, the propensity score literature recommends combining propensity score matching with regression (including the same confounders in both models) in order to best control for confounding. [35] This type of analysis has the advantage of being doubly robust,

because estimates are consistent if either the propensity score model is correct or if the outcome model is correct. [35] Gamma regression models with a log link were used for the point-in-time cortisol measures. Gamma regression is appropriate for positive, skewed data, as was the case for these cortisol data, and gamma regression coefficients have a straightforward multiplicative interpretation. [20] The exponentiated neighborhood regression coefficient from each of the pre- and post-interview cortisol models estimates the conditional multiplicative effect of living in a disadvantaged neighborhood on point-in-time cortisol level. Linear regression was used to model cortisol slope. The neighborhood regression coefficient for this model estimates the conditional additive effect of living in a disadvantaged neighborhood on cortisol slope.

We ran one unadjusted and three adjusted models for each cortisol outcome. Adjusted Model 1 contained potential confounding variables (listed in Figure 3.1) that we believe preceded the adolescent's neighborhood residence. Adjusted Model 2 added variables that could be thought of as confounders or mediators: household income and current and historical parental employment. Adjusted Model 3 added hypothesized mediators of the neighborhood disadvantage-cortisol association: sleep, BMI, and small for gestational age. Results for each outcome model were combined across the ten imputed datasets using Rubin's combining rules. [62] All analyses were performed using the R statistical language (version 2.15.1).

3.4 Results

Table 3.3 describes the demographic characteristics of adolescents included in the final matched dataset by neighborhood disadvantage status. The propensity score

matching procedure described above resulted in adolescents having similar demographic characteristics between disadvantaged and non-disadvantaged neighborhoods. The mean age was just under 15 years, and 48% were female. Hispanic, African American, and White racial/ethnic categories were well represented (31%, 37%, and 28%, respectively), as were urban, suburban and rural areas (32%, 41%, and 27%, respectively). In this matched dataset, 62% of adolescents were from the Southern region of the U.S., 51% had mothers with a high school education, and 41% were interviewed in the summer.

We estimated the conditional expected ratios of pre- and post-interview cortisol levels and conditional expected differences in cortisol rate of change between those living in disadvantaged versus non-disadvantaged neighborhoods running regression-adjusted models on matched datasets balanced on covariates of interest. As described in the Methods section, cortisol measures were limited to those occurring after the cortisol awakening response, so inference is limited to the late decline portion of cortisol's circadian rhythm. Unadjusted model results are included for comparison.

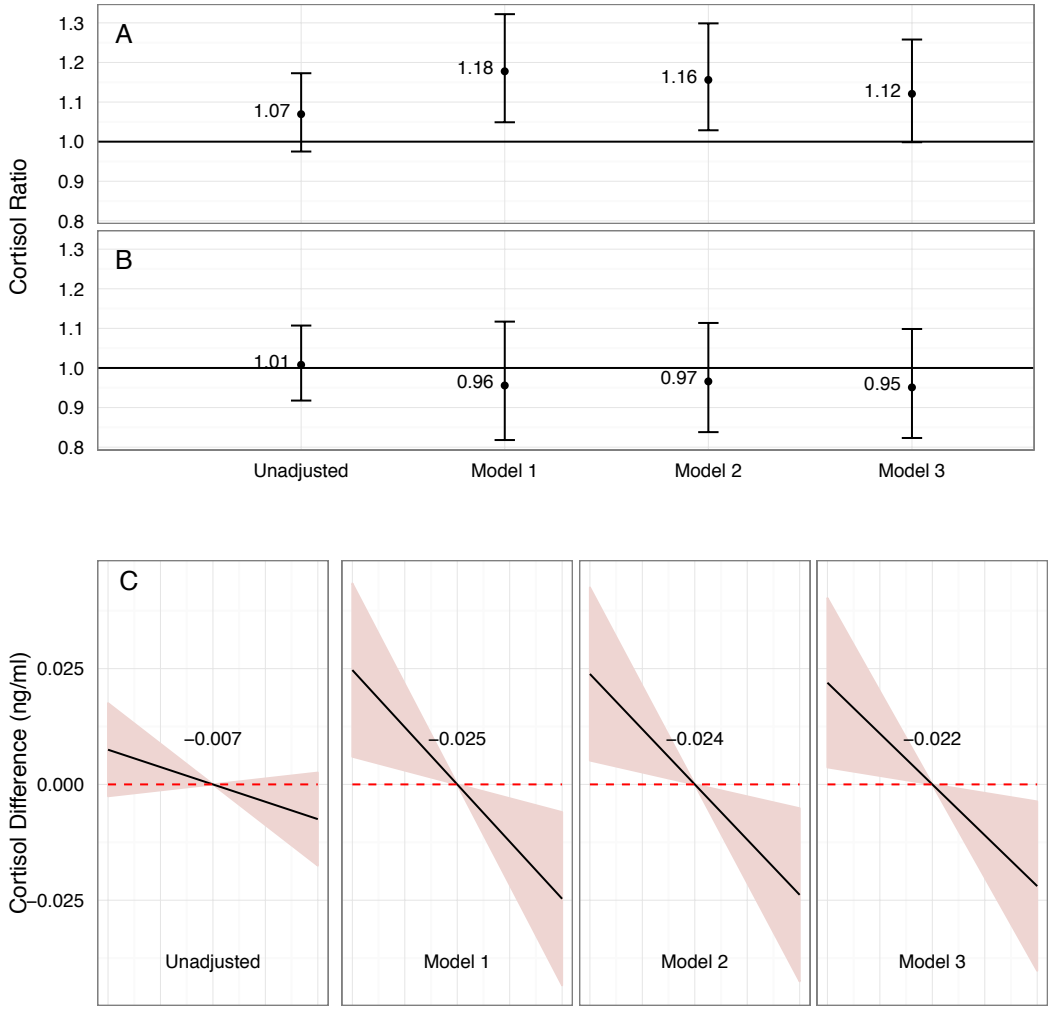
Figure 3.2 shows the associations between neighborhood disadvantage and (A) pre-interview cortisol, (B) post-interview cortisol, and (C) cortisol rate of decline for the unadjusted and adjusted models. In the unadjusted model, adolescents living in disadvantaged neighborhoods had 7% higher pre-interview cortisol levels than those in non-disadvantaged neighborhoods (95% CI, 1.00, 1.15), 0.8% higher post-interview cortisol levels (95% CI, 0.93, 1.09), and 0.75 (95% CI, -0.11, 1.60) $\times$10-2 ng/mL/hr steeper rates of decline (Figure 3.2).

Model 1 adjusts for covariates listed in Figure 3.1. In this model, adolescents in disadvantaged neighborhoods had 19% higher (95% CI, 1.05, 1.34) pre-interview cortisol levels than those in non-disadvantaged neighborhoods, 4% lower (95% CI, 0.85,

1.07) post-interview cortisol levels, and 2.47 (95% CI, 0.81, 4.13) ×10-2 ng/mL/hr steeper rates of decline (Figure 3.2). Model 2 adjusts for the covariates in Model 1 and also adjusts for household income and current and previous parental employment. Effect sizes attenuated slightly in this model, but neighborhood disadvantage residence remained associated with higher pre-interview cortisol levels (1.15, 95% CI: 1.00, 1.31) and steeper rates of decline (2.38, 95% CI: 0.61, 4.14).

Small for gestational age, BMI, and sleep are all potential mediators of the neighborhood disadvantage-cortisol relationship. Adjusted Model 3 added these mediating variables to the covariates included in Adjusted Model 2. The effect sizes attenuated slightly comparing Model 2 to Model 3 (Figure 3.2). The association between neighborhood disadvantage residence and higher pre-interview cortisol levels was no longer significant at the 95% confidence level (1.13, 95% CI: 0.99, 1.28), but the association between neighborhood disadvantage and steeper rates of decline remained statistically significant (2.20, 95% CI: 0.52, 3.87).

Figure 3.2: Conditional expected ratios of cortisol levels and conditional expected differences in cortisol slope during the late decline period comparing adolescents living in disadvantaged versus non-disadvantaged neighborhoods. Models were matched on and regression-adjusted for covariates listed in Figure 3.1. Row A: Ratios of point-in-time pre-interview cortisol levels. Error bars represent 95% CI for the mean. Row B: Ratios of point-in-time post-interview cortisol levels. Error bars represent 95% CI for the mean. Row C: Differences in cortisol slope. Shaded areas represent 95% CI for the mean.



We also examined the potential explanatory power of each of these mediators by individually adding them to Adjusted Model 2. Neither the addition of BMI nor small for gestational age appreciably changed the size of the neighborhood disadvan-

tage parameter for any of the cortisol outcomes, and inferences remained the same as in Adjusted Model 2. Sleep variables included the time the adolescent went to bed on weekday nights, the time the adolescent went to bed on weekend nights, hours of sleep on weekdays, and hours of sleep on weekends. The addition of sleep variables attenuated the association between neighborhood disadvantage residence and pre-interview cortisol levels by about 9% and rendered it non-significant. Adding sleep variables also attenuated the association between neighborhood disadvantage and cortisol rate by about 6%, but it remained statistically significant.

3.5 Discussion

In a large, U.S. population-based sample, we found that adolescents living in disadvantaged neighborhoods had higher pre-interview cortisol levels and steeper declines in cortisol over the course of the interview, perhaps reflecting heightened reactivity to and recovery from the novel stimulus of the interview. There were no differences in post-interview cortisol levels. These results add to the nascent body of literature that links neighborhood context and stress in adolescents.

This study addressed several gaps in the literature. Previous research has been hampered by small sample sizes and ethnic and geographic homogeneity. We addressed this by using a large, population-based, and ethnically diverse sample of adolescents. In addition, our use of propensity score matching methods coupled with regression adjustment addressed confounding stemming from non-random neighborhood residence.

According to McEwen, there are two key dimensions of the stress response sys-

tem: (1) reactivity to an acute stressor and (2) unprovoked, chronic levels of stress hormones. [51] Both could become dysregulated as a result of chronic stress caused by residence in a disadvantaged neighborhood. To measure these dimensions precisely requires controlled laboratory conditions that hold time, day of the week, and certain environmental variables (e.g., light, temperature, noise) and behaviors (e.g., level of physical activity, smoking, drinking) constant and protocols that administer either a standardized acute stressor (e.g., a Trier stress test) or rest period. [31] Because such a protocol was not possible, the cortisol outcomes used in this analysis do not map onto specific HPA axis dimensions.

As described previously, the pre-interview sample may reflect the adolescent's anticipation of the new experience of being interviewed by a stranger as part of a survey of mental health (which may provoke an HPA response in some) and may be influenced by whatever activity the adolescent was engaged in prior to the interview. The post-interview sample may reflect both a natural decline as per cortisol's diurnal rhythm and perhaps a more restful state, as the adolescent has been seated for an average of 2 $\frac{1}{2}$ hours and has finished answering questions that were designed not to be stressful. The rate of decline in cortisol levels between pre- and post-interview may reflect recovery from both the anticipation of the novel experience of the interview and activity prior to the interview.

Prior studies of the relationship between resting cortisol and neighborhood- and individual-level SES have yielded mixed results, possibly due to effect heterogeneity.. [13, 19, 29, 36] Thus, while we hypothesized higher post-interview cortisol levels in adolescents who resided in disadvantaged neighborhoods, our null results are not inconsistent with prior literature. We could have obtained null results either because there is no association between neighborhood disadvantage and post-interview cor-

tisol levels or because there is an association but our results are biased or because there is effect heterogeneity that we have not accounted for. No association between post-interview levels and neighborhood disadvantage suggests that there is no difference in cortisol levels after sitting for a 2 $\frac{1}{2}$ hour interview between adolescents in disadvantaged versus non-disadvantaged neighborhoods.

There has been little research on the relationship between cortisol response to novel situations and neighborhood- and individual-level SES. However, animal research demonstrates that rats stressed when young (e.g., by a lack of maternal licking/grooming) are more "neophobic," with heightened, less controlled stress responses to novel situations that persist into adulthood. [44, 45, 52] In human children, Gutteling et al. found that children of anxious mothers had higher cortisol levels on the first day of kindergarten than children of non-anxious mothers, [27] and Hackman et al. found in a sample of African American adolescents that those in disadvantaged neighborhoods had greater cortisol reactivity to a stress test than those in non-disadvantaged neighborhoods. [28]

Our findings are congruent with this prior research. However, we recognize that the steeper slope found for adolescents living in disadvantaged neighborhoods may seem counterintuitive. Higher pre-interview cortisol levels and steeper slope may imply a hypervigilant HPA axis, but one that is still resilient and healthy. As a cross-sectional survey of adolescents, this study captures a particular developmental window. Over time, repeated or prolonged hypervigilant responses may succumb to allostatic load, eventually resulting in desensitization of the HPA axis. [22, 51]

3.5.1 Sensitivity Analyses

We performed four sensitivity analyses. First, we re-ran the analysis (1) excluding current smokers but including current drug users and (2) including both current drug users and current smokers. This tested the sensitivity of our results to the decision to exclude adolescents who are current smokers and drug users. (Such exclusion criteria are common, as these substances may affect cortisol levels and responsiveness (e.g., [34,41]).) With each inclusion, the effect sizes attenuate (see Table 3.4). This is expected as we are including individuals whose cortisol levels may be artificially influenced and therefore may not be at-risk for being influenced by neighborhood sources of stress. When we exclude smokers only, neighborhood disadvantage remains associated with cortisol rate of decline, neighborhood disadvantage and post-interview cortisol levels remain unassociated, and the association between neighborhood disadvantage and pre-interview cortisol levels remains significant in Adjusted Model 1 but not in Adjusted Model 2. When we include both current drug users and smokers, there is no longer a statistically significant association between neighborhood disadvantage and pre-interview cortisol levels in Adjusted Models 1 or 2. The association between neighborhood disadvantage and cortisol rate remains significant at the 95% confidence level in Adjusted Model 1, but not in Model 2.

Table 3.4: Conditional expected ratios in cortisol levels and conditional expected differences in slope (ng/mL/hr $\times 10^{-2}$) during the late decline portion of cortisol's circadian rhythm comparing adolescents living in disadvantaged versus non-disadvantaged neighborhoods under different exclusion criteria.

Model	**Pre**	**Post**	**Rate**
	Mean (95% CI)	Mean (95% CI)	Mean (95% CI)
Current smokers and drug users excluded			
Unadjusted	1.07 (1.00, 1.15)	1.01 (0.93, 1.09)	-0.75 (-1.60, 0.11)
Model 1	1.19 (1.05, 1.34)	0.96 (0.85, 1.07)	-2.47 (-4.13, -0.81)
Model 2	1.15 (1.00, 1.31)	0.96 (0.86, 1.07)	-2.38 (-4.14, -0.61)
Model 3	1.13 (0.99, 1.28)	0.95 (0.85, 1.06)	-2.20 (-3.87, -0.52)
Current smokers excluded, current drug users included			
Unadjusted	1.07 (1.00, 1.15)	1.00 (0.92, 1.09)	-0.78 (-1.59, 0.03)
Model 1	1.16 (1.01, 1.33)	0.94 (0.84, 1.07)	-2.30 (-3.89, -0.72)
Model 2	1.13 (0.97, 1.32)	0.95 (0.85, 1.06)	-2.21 (-3.95, -0.47)
Model 3	1.11 (0.95, 1.28)	0.94 (0.84, 1.05)	-2.03 (-3.68, -0.38)
Current smokers and drug users included			
Unadjusted	1.06 (1.00, 1.13)	1.02 (0.94, 1.11)	-0.56 (-1.29, 0.18)
Model 1	1.11 (0.96, 1.27)	0.96 (0.84, 1.10)	-1.77 (-3.36, -0.18)
Model 2	1.08 (0.92, 1.26)	0.97 (0.86, 1.10)	-1.62 (-3.34, 0.10)
Model 3	1.06 (0.92, 1.23)	0.96 (0.86, 1.08)	-1.44 (-3.14, 0.27)

Second, we re-ran the analysis excluding adolescents who may have experienced severe adversity during childhood. We operationalized this as excluding individuals who reported ever being physically abused by a parent (including being beaten

up, choked, burned, kicked, punched, bitten, or threatened with a knife or gun) or reported three or more of the following: parent suicide attempt, parent suicide completion, parent alcohol abuse, parent drug abuse, parent arrest or imprisonment. The literature indicates that timing and severity of exposure to adverse conditions alter HPA axis dynamics in adolescents. [7] Adolescents with moderate levels of childhood adversity may have a hypervigilant but resilient stress response, whereas adolescents with severe, persistent adversity may undergo a degree of desensitization leading to blunted cortisol response. [27, 48, 58] Failure to distinguish between these two subgroups likely results in a partial washout effect. A greater proportion of adolescents with possible severe adversity lived in disadvantaged neighborhoods (see Table 3.3), and we would expect from prior research that they would be more likely to have blunted HPA axis reactivity. We expected that excluding these adolescents would strengthen the associations between living in a disadvantaged neighborhood and pre-interview cortisol levels and cortisol slope. However, effect estimates changed very little and all inferences remained the same, possibly because the disadvantaged and non-disadvantaged neighborhood groups were not much different in terms of possible severe adversity after matching (see Table 3.3).

Third, we assessed the sensitivity of our results to our decision to use the first tertile to define neighborhoods as disadvantaged. We repeated our analysis using the 25th, 20th, 15th, 10th, and 5th percentile as cut-points (results not shown, but available from the first author upon request). The association between neighborhood disadvantage and cortisol slope remained statistically significant for Adjusted Model 1 for all cut-points except the 5th percentile. In this case, the association was no longer significant, possibly because the sample size of the matched dataset was greatly reduced in this case (from N=867 to N=148 for the first imputation). For each

alternative cut-point, the parameter estimate of neighborhood disadvantage on pre-interview cortisol levels remained similar but was no longer significant at the 95% confidence level, possibly also due in part to a reduction in sample size.

We found the association between pre-interview cortisol and neighborhood disadvantage to be sensitive to the current smoker and drug user exclusion criteria in the first sensitivity analysis and sensitive to different cut-points of neighborhood disadvantage in the third sensitivity analysis. In contrast, the association between cortisol rate of change and neighborhood disadvantage was robust. We examined the robustness of this association to an unobserved confounder in a fourth sensitivity analysis. Assessing sensitivity to an unobserved confounder is important because although the above methods aim to approximate the comparability of exposed and unexposed groups found in a randomized control trial, the exposed and unexposed groups could differ by an unobserved confounder. We used a bias equation described in VanderWeele and Arah for average effect differences for those with versus without the exposure, conditional on a vector of confounding variables, $\mathbf{X}$. [69] We made the following three simplifying assumptions as discussed by VanderWeele and Arah: (1) the association between the outcome and unobserved confounder is consistent across levels of exposure and observed covariates; (2) the unobserved confounder is binary; and (3) the difference in the prevalence of the unobserved confounder in the exposed versus unexposed is constant across levels of the covariates. Let A denote the exposure received by an adolescent. In this study, $a1$ represents residence in a disadvantaged neighborhood and $a0$ represents residence in a non-disadvantaged neighborhood. Let Y denote the observed outcome, cortisol rate of change. Let $\mathbf{X}$ denote the observed covariates, and let U denote the unobserved binary confounder.

Under the above three simplifying assumptions, the bias of the estimated effect

is:

$$d(x) = \delta\gamma$$

where

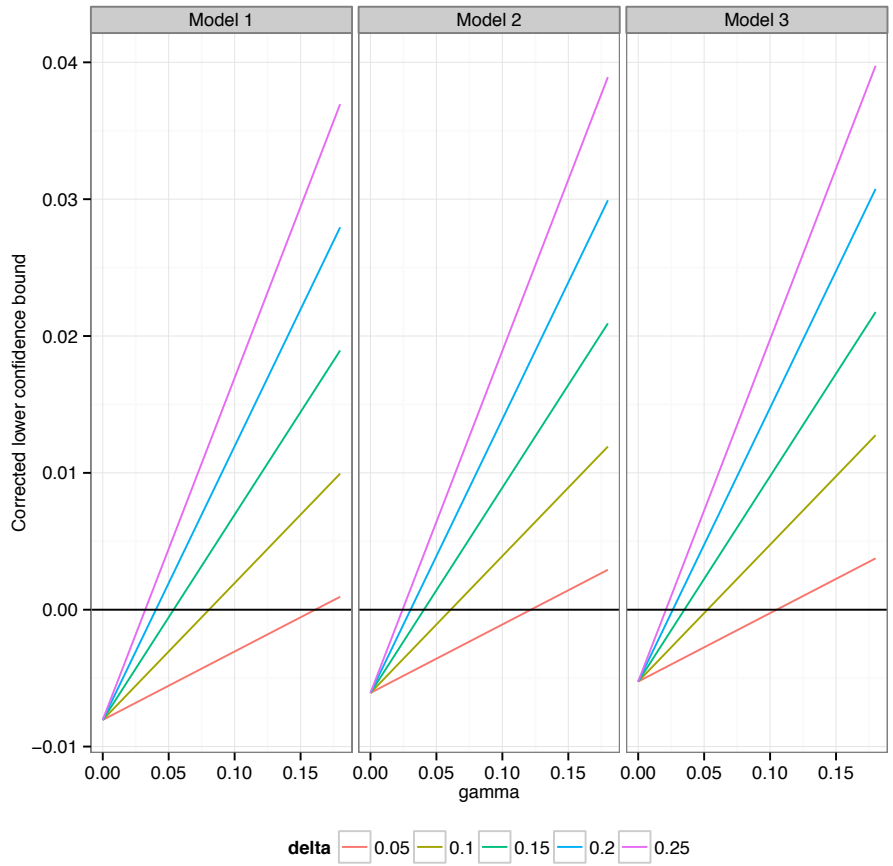
$$\delta = P(U = 1|a1, \mathbf{X}) - P(U = 1|a0, \mathbf{X}) \text{ and } \gamma = E(Y|a, \mathbf{X}, U = 1) - E(Y|a, \mathbf{X}, U = 0)$$

We calculated the bias and resulting corrected lower 95% confidence bound across an array of input parameter values. We allowed γ to range from 0 to 0.18. We allowed δ, the difference in prevalence of the unobserved confounder between exposed and unexposed groups, to range from 5 to 25%.

Figure 3.3, below, plots the corrected lower 95% CI bound (on the y-axis) against the value of γ (on the x-axis). The different colored curves in each subplot show the relationship for different values of δ. Each of the three subplots corresponds to the three adjusted models used in the study (see Figure 3.2). For Adjusted Model 1 and setting δ=0.2, γ would have to be at least 0.04 to render our estimate of the average difference in cortisol slope between disadvantaged and nondisadvantaged neighborhoods nonsignificant. Put another way, the presence of an unobserved confounder would have to change the average effect by 62% to render the effect non-significant.

Including possible mediators in the model makes the effect estimate more sensitive to an unobserved confounder. For example, in Adjusted Models 2 and 3, γ would need to be greater than 0.03 and 0.026 (an increase of 26% and 18% over the average effect size), respectively.

Figure 3.3: Corrected lower 95% confidence bound by values of δ and γ



3.5.2 Strengths and Limitations

In terms of this study, internal validity is compromised if the adolescents who live in disadvantaged neighborhoods are still not comparable (also called exchangeable) with those who live in non-disadvantaged neighborhoods after controlling for covariates. Our analytic approach addresses this challenge in several ways not often used in studies of this kind. First, the NCS-A collected extensive information from adolescents and their parents that we were able to use to control for confounding. Second,

our matching analytic design allowed us to explicitly assess covariate balance between the disadvantaged neighborhood group and non-disadvantaged neighborhood group, and perform the analysis on a subset of the data in which the two groups are balanced across our list of potential covariates. Third, using the above matching methods in conjunction with regression can be thought of as a doubly-robust approach, meaning that if either the propensity-score exposure model or outcome regression model is correct, then the estimates are unbiased in expectation. [35]

Our results are subject to several limitations. The summary measure of neighborhood disadvantage comprises a set of indicators measuring different components of neighborhood disadvantage, like income, assets, housing value, education, and employment, which should contain less random measurement error than any one indicator alone. However, neighborhood disadvantage is still measured with error. First, the summary measure does not capture all the characteristics that define the latent construct of neighborhood disadvantage. Second, we use Census tracts as a proxy for neighborhood; neighborhood boundaries identified by residents will likely not overlap completely with Census tract boundaries.

Another limitation is that cortisol is just one of many biomarkers of HPA axis activity. Free, unbound cortisol can be obtained from saliva samples whereas total—bound and unbound—cortisol can be obtained from serum samples. Unbound cortisol is likely the more relevant proxy, because it is thought to be the only component of cortisol to reach the “target tissue and elicit glucocorticoid effects". [41] Consistent with this idea, saliva cortisol has been found to better measure adrenal cortical function [70] and HPA axis activity, [25] although saliva and serum cortisol are highly correlated. [72] In addition, measuring cortisol through saliva samples is non-invasive and does not induce stress—a strength because of cortisol’s sensitivity to stress. [41]

In addition, the variability of cortisol presents a major challenge. Each person has her/his own activation thresholds and unique diurnal pattern. Within person, cortisol levels depend on time of day, day of the week, activities, diet, amount of sunshine, etc. Other epidemiologic studies incorporating cortisol measures in adult samples have sought to address variability within the limitation of a non-laboratory environment by modeling within- versus between-person variability of the natural diurnal rhythm of cortisol levels by collecting multiple measurements per person per day for multiple days. However, this collection scheme has been shown to be infeasible with adolescents, [30] so we instead applied exclusion criteria and restrictive matching methods to move the analysis back to a scenario that holds three of the driving sources of variability either constant or close to constant and limits the analysis to a sample whose cortisol levels are not artificially influenced by prescription or non-prescription drugs, hormones, or tobacco. This may provide an initial way forward for the practical researcher wanting to make use of cortisol measurements in large, epidemiologic studies.

Another limitation is that we cannot infer clinical significance from our results. Dysregulation of the stress response system may increase risk of mental, cardiovascular, immunologic, and metabolic disorders. [3, 66] Dysregulation of the HPA axis has also been implicated in unhealthy eating behaviors and obesity. [51, 65] Future research should examine dose-response relationships and possible threshold effects between cortisol measures and these health outcomes. Although the associations we found were small, seemingly small differences that persist over many years may ultimately have profound effects on neurocircuitry and glucose regulation.

3.5.3 Conclusion

In conclusion, we found evidence of a heightened, yet resilient, cortisol response to a novel interview situation among adolescents living in disadvantaged versus non-disadvantaged neighborhoods in an ethnically and geographically diverse sample. This extends previous animal and laboratory-based research and bolsters the evidence base suggesting that place may influence the stress response system. Implications of such a conclusion argue for policies designed to improve the safety and built environment of the U.S.'s most disadvantaged neighborhoods. More research is needed to further understand individual and contextual determinants as well as to inform specific policy targets.

Chapter 4

Estimating population treatment effects from a survey sub-sample

This chapter's research has been submitted and is under review for publication. [26] All code associated with this chapter can be found here: https://github.com/cherrygarcia/Aim3.

4.1 Abstract

We consider the problem of estimating an average treatment effect for a target population from a survey sub-sample. Our motivating example is generalizing a treatment effect estimated in a sub-sample of the National Comorbidity Survey Replication Adolescent Supplement to the population of U.S. adolescents. To address this problem, we evaluate easy-to-implement methods that account for both non-random treatment assignment and a non-random two-stage selection mechanism. We compare the performance of a Horvitz-Thompson estimator using inverse probability weighting (IPW) and two double robust estimators in a variety of scenarios. We demonstrate that the two double robust estimators outperform IPW in terms of percent bias, variance, and mean squared error, even under misspecification of one of the treatment,

selection, or outcome models. Moreover, the double robust estimators are easy to implement, providing an attractive alternative to IPW for applied epidemiologic researchers. We then demonstrate how to apply these estimators to our motivating example.

4.2 Introduction

Population-based cohorts and nationally representative surveys lend external validity to a study: they allow inferences to be made about the target population of interest. In contrast, inferences drawn from studies that use non-representative samples may be valid for the study sample but may not generalize. External validity (also known as transportability [19]) of population-based cohorts and surveys is threatened when estimation is performed on a non-random sub-sample. Sub-sample effect estimates may not generalize to the population if selection probabilities depend on effect modifiers and if sub-sample sampling weights are not calculated [4, 30]. In this paper, we compare practical estimators of the population average treatment effect. These estimators simultaneously account for non-randomized treatment assignment and sub-sample selection from a population-based cohort, thereby addressing internal and external validity.

This paper was motivated by the problem of generalizing a treatment effect estimated in a sub-sample created by a two-stage selection process. In the first stage, adolescents were selected into a nationally representative survey assessing U.S. adolescent mental health, the National Comorbidity Survey Replication Adolescent Supplement (NCS-A) [16]. In the second stage, a sub-sample of these participants had biomarker data measured. Our interest is in estimating the effect of a non-randomized

treatment, residence in a disadvantaged neighborhood, on cortisol slope. Our scenario is different from the missing data pattern generally considered in the causal inference literature because we do not observe any data for individuals not in the survey. Our goal is to harness the available data and the nationally representative sample to generalize our results to the U.S. population of adolescents. This requires accounting for possible confounding due to the non-randomized treatment assignment and possible lack of external validity due to the two-stage selection mechanism.

Previous research has suggested and evaluated methods for generalizing results from randomized trials to target populations [4, 30], but there is little written extending this to observational studies. Double robust methods, which are consistent (converge to the true population average effect as sample size goes to infinity) under certain types of model misspecification, have been used to adjust for non-random treatment assignment and/or non-random selection, right-censoring, or missing data [1, 2, 7, 9, 13, 21, 24, 28, 31, 32]. However, implementation of these estimators can be challenging. This may contribute to the continued popularity of the simpler Horvitz-Thompson inverse-probability weighted (IPW) estimators despite efficiency concerns [23]. A recent advance is that targeted maximum likelihood estimation (TMLE) was implemented in standard statistical software [7], thereby facilitating its accessibility. However, we know of no literature implementing a TMLE in the context of survey data with weights. Furthermore, while a discussion of these methods is taking place in the biostatistics literature, it has yet to receive much attention in the epidemiology literature.

We first present results from a simulation study comparing performance of different estimators under various data generating distributions and model misspecifications. We compare three estimators: IPW; a double robust, weighted least squares estimator

(DRWLS); and a TMLE [8,33]. We then demonstrate how to apply these methods to the motivating example: using data from a sub-sample of the NCS-A to estimate the effect of residence in a disadvantaged neighborhood on cortisol slope in the population of U.S. adolescents. We aim to provide practical guidance on how to generalize average effect estimates from a survey sub-sample to a target population in the presence of measured confounders, effect heterogeneity, and non-random sub-sample selection.

4.3 Description of Methods

We consider a scenario in which individuals are selected into a survey with known probabilities. Treatment information and covariates are fully observed for all participants selected into the survey, but outcome data are only available for a subset of the survey sample. Let:

$\mathbf{W}$ = vector of baseline covariates.

A = binary (0/1) variable indicating treatment.

Δ_{svy} = binary (0/1) variable indicating selection into survey sample.

Δ_{sub} = binary (0/1) variable indicating selection into sub-sample.

Y = continuous outcome of interest.

In the language of potential outcomes, Y_{1i} is the outcome for individual i if treatment $A = 1$ would be assigned; similarly, Y_{0i} is the outcome for individual i if treatment $A = 0$ would be assigned. The difference in these potential outcomes is the individual treatment effect. Our estimand of interest is the average treatment effect,

$E(Y_1 - Y_0)$ [25], with the expectation taken across the target population. Under certain assumptions (strongly ignorable treatment assignment, the stable-unit treatment value assumption, well-defined intervention, known survey probabilities), this estimand is identifiable, as shown in Equation 4.1 below. Denote the true regression function of the outcome on the covariates among individuals in the sub-sample with treatment level $A = a$ by $\mu(a, \mathbf{W})$.

$$E(Y_1 - Y_0) = \frac{E\{\pi(W_1)^{-1}(\mu(1, \mathbf{W}) - \mu(0, \mathbf{W}))|\Delta_{svy} = 1\}}{E\{\pi(W_1)^{-1}|\Delta_{svy} = 1\}}, \tag{4.1}$$

where the probability of survey selection is known and defined as $P(\Delta_{svy} = 1|\mathbf{W}) = \pi(W_1)$.

We now show the identifiability result from Equation 4.1 under the assumption of no unmeasured confounders and positivity of the probabilities $P(A = a, \Delta_{sub} = 1, \Delta_{svy} = 1|\mathbf{W})$ for $a = 0, 1$.

$$\begin{aligned}
E(Y_1 - Y_0) &= E[E(Y|A = 1, \Delta_{sub} = 1, \Delta_{svy} = 1, \mathbf{W}) \\
&\quad -E(Y|A = 0, \Delta_{sub} = 1, \Delta_{svy} = 1, \mathbf{W})] \\
\text{let } \mu(1, \mathbf{W}) &= E(Y|A = 1, \Delta_{sub} = 1, \Delta_{svy} = 1, \mathbf{W}) \\
\text{and let } \mu(0, \mathbf{W}) &= E(Y|A = 0, \Delta_{sub} = 1, \Delta_{svy} = 1, \mathbf{W}) \\
E(Y_1 - Y_0) &= E[(\mu(1, \mathbf{W}) - \mu(0, \mathbf{W}))]
\end{aligned}$$

To make the above estimable from an iid sample of O, we write:

$$\begin{aligned} E(Y_1 - Y_0) &= E_{W1}\{E[(\mu(1,\mathbf{W}) - \mu(0,\mathbf{W}))|W_1, \Delta_{svy} = 1]\} \\ &= \frac{E\{\pi(W_1)^{-1}(\mu(1,\mathbf{W}) - \mu(0,\mathbf{W}))|\Delta_{svy} = 1\}}{E\{\pi(W_1)^{-1}|\Delta_{svy} = 1\}} \\ &= \frac{\frac{1}{n}\sum_{i=1}^{n} \pi_i(W_1)^{-1}(\mu_i(1,\mathbf{W}) - \mu_i(0,\mathbf{W}))}{\frac{1}{n}\sum_{i=1}^{n} \pi_i(W_1)^{-1}}. \end{aligned}$$

We now detail the three methods we compare. The code used for each method is provided here:

https://github.com/cherrygarcia/Aim3/Functions.

4.3.1 IPW

The IPW estimator uses inverse probability of treatment and selection weights that are obtained by multiplying inverse probability of survey selection weights, inverse probability of treatment weights, and inverse probability of sub-sample selection weights.

Inverse probability of treatment weights are defined as:

$$\begin{aligned} w^{A=1|\Delta_{svy}=1} &= \frac{I(A=1)}{P(A=1)|\Delta_{svy}=1,\mathbf{W})} \\ w^{A=0|\Delta_{svy}=1} &= \frac{I(A=0)}{P(A=0)|\Delta_{svy}=1,\mathbf{W})} \end{aligned}$$

Inverse probability of sub-sample selection weights are defined as:

$$w^{\Delta_{sub}=1|A=a,\Delta_{svy}=1} = \frac{I(\Delta_{sub}=1)}{P(\Delta_{sub}=1|A=a,\Delta_{svy}=1,\mathbf{W})}$$

These two weights are multiplied with the survey weights to give the inverse probability of treatment and selection weights:

$$
\begin{aligned}
w^{A=1,\Delta_{svy}=1,\Delta_{sub}=1} &= \frac{I(\Delta_{svy}=1)}{P(\Delta_{svy}=1|\mathbf{W})} \times \frac{I(A=1)}{P(A=1)|\Delta_{svy}=1,\mathbf{W})} \\
&\times \frac{I(\Delta_{sub}=1)}{P(\Delta_{sub}=1|A=1,\Delta_{svy}=1,\mathbf{W})} \\
&= \frac{I(A=1,\Delta_{sub}=1,\Delta_{svy}=1)}{P(A=1,\Delta_{sub}=1,\Delta_{svy}=1|\mathbf{W})} \\
w^{A=0,\Delta_{svy}=1,\Delta_{sub}=1} &= \frac{I(\Delta_{svy}=1)}{P(\Delta_{svy}=1|\mathbf{W})} \times \frac{I(A=0)}{P(A=0)|\Delta_{svy}=1,\mathbf{W})} \\
&\times \frac{I(\Delta_{svy}=1)}{P(\Delta_{sub}=1|A=0,\Delta_{svy}=1,\mathbf{W})} \\
&= \frac{I(A=0,\Delta_{sub}=1,\Delta_{svy}=1)}{P(A=0,\Delta_{sub}=1,\Delta_{svy}=1|\mathbf{W})}
\end{aligned}
\tag{4.2}
$$

For the IPW estimator, the inverse probability of treatment and selection weights are multiplied by the outcome Y and averaged over the r individuals in the sub-sample to estimate the population average treatment effect: $\frac{1}{\sum_{i=1}^{r} w_i^{A=a,\Delta_{svy=1},\Delta_{sub}=1}} \sum_{i=1}^{r} w_i(2A_i - 1)Y_i$.

4.3.2 TMLE

For the TMLE estimator, we modify the implementation available in the tmle R package [7] as shown here: https://github.com/cherrygarcia/Aim3/Functions. We also provide R code to implement this estimator without the tmle package here: https://github.com/cherrygarcia/Aim3/tmleFuncWithoutRpackage.

Below we summarize the main steps involved.

1. Obtain predicted values $\hat{Y}^0$ of the outcome conditional on the treatment and covariates using a linear regression of Y as a function of A and $\mathbf{W}$.

2. Estimate coefficients for the vector of clever covariates $(w^{A=0,\Delta_{sub}=1,\Delta_{svy}=1}, w^{A=1,\Delta_{sub}=1,\Delta_{svy}=1})$ via regression with Y as the outcome and using $\hat{Y}^0$ as an offset.

3. Update $\hat{Y}^0$ by the estimated coefficients from Step 2 multiplied by the vector of clever covariates. That is, $\hat{\mu}^1(A,\mathbf{W}) = logit^{-1}\left(logit(\hat{\mu}^0(A,\mathbf{W}))+\hat{\epsilon}\frac{(2A-1)\Delta_{sub}}{\pi(W_1)g(A,\mathbf{W})}\right)$, where $g(A,\mathbf{W}) = \hat{P}(A=a,\Delta_{sub}=1|\mathbf{W})$. Extract both predicted counterfactual outcome values (setting $A=1$ and $A=0$) for each individual in the survey sample using G-computation. Weighting the predicted individual treatment effects by the survey weights, compute the population average effect.

   The TMLE estimator is the solution to the efficient influence function:

$$D(\hat{O}) = 0 = \frac{1}{\pi(W_1)}\left[\frac{(2A-1)\Delta_{sub}}{g(A,\mathbf{W})}\left(Y_i - \mu^1(A,\mathbf{W})\right) + \mu(1,\mathbf{W}) - \mu(0,\mathbf{W}) - \hat{\psi}\right]$$

For readers unfamiliar with G-computation, Snowden et al (2011) provide an introduction. [29] Briefly, G-computation uses the marginal distribution of covariates in a standardization procedure. This can be thought of as an extension of standardizing mortality rates by the age distribution in a standard population—a common epidemiologic practice. One fits an outcome model in the observed sample and then applies the model to the distribution of covariates in the standard population to predict the counterfactual outcomes for each individual: $\hat{Y}_1|A=1,\mathbf{W}$ and $\hat{Y}_0|A=0,\mathbf{W}$.

4.3.3 DRWLS

The DRWLS estimator combines weighted regression with G-computation. This estimator was first suggested by Marshall Joffe [23] and has been previously evaluated.

[9, 23] It is constructed in the following steps:

1. Use $w^{A=a,\Delta_{sub}=1,\Delta_{svy}=1}$ as weights in a weighted least squares (WLS) regression outcome model. Using G-computation, predict counterfactual outcomes from WLS, standardized to the survey sample.

2. Use the counterfactual outcomes to estimate the average effect in the survey sample.

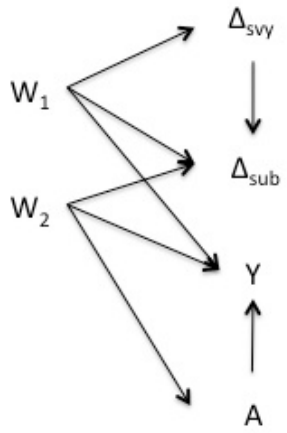
3. Weight this estimate by the survey weights to estimate the averag effect in the population.

4.4 Simulation Study

4.4.1 Overview and set-up

We consider a simplified case with two continuous covariates: $\mathbf{W} = [W_1, W_2]'$. Let observed data $O = (\Delta_{svy} = 1, \mathbf{W}, A, \Delta_{sub}, \Delta_{sub}Y)$. We assume Δ_{svy} probabilities are known, there are no unobserved confounders, and no additional intermediate variables exist between A and Y. Figure 4.1 depicts the data-generating mechanism.

Figure 4.1: Data generating mechanism.



For each of 1,000 simulations, we generate 100,000 simulated population members, each with a complete data vector. First, we generate covariates from independent normal distributions: $W_1 \sim N(0,1), W_2 \sim N(2,1)$. Second, we generate an indicator of survey selection from a Bernoulli distribution with probability: $P(\Delta_{svy} = 1|\mathbf{W}) = Logit^{-1}(-2.3 + 0.7W_1)$. Approximately 10,000 population members (10%) are retained in the survey. Third, we generate an exposure variable from a Bernoulli distribution with probability: $P(A = 1|\mathbf{W}) = Logit^{-1}(-2.5 + 0.4W_2 + 0.2W_2^2)$. Approximately one-third of the population is exposed ($A = 1$). Fourth, we generate an indicator of sub-sample selection from a Bernoulli distribution with probability: $P(\Delta_{sub} = 1|\Delta_{svy} = 1, \mathbf{W}) = Logit^{-1}(-2 + 0.2W_1 + 0.2W_1^2 + 0.7W_2)$. Approximately 5,000 (50% of survey sample, 5% of the population) are retained in the subsample. Finally, we generate one continuous outcome variable for each of two scenarios—linear effect heterogeneity and non-linear effect heterogeneity. In Scenario 1, $Y = -3 + W_2 + W_2^2 + 3A + AW_1 + \epsilon, \epsilon \sim N(0,2)$. In Scenario 2, $Y = 2W_2 + 2A + 2AW_1 + 2AW_1^2 + \epsilon, \epsilon \sim N(0,2)$. Under both scenarios, the true

population average effect is 3.

Figure 4.2 provides a diagram of the complete and observed data. The complete data vector for each simulated population member consists of $W_1, \Delta_{svy} = 1, W_2, A, \Delta_{sub} = 1$ and counterfactual outcomes Y_0 and Y_1. We do not observe these counterfactual outcomes directly. Under the causal inference framework we are using, it is assumed that the observed data generating process consists of several steps, starting with the full data and ending in what we actually observe in our dataset. First, selection into the survey (Δ_{svy}) is determined, where the probability of selection depends on W_1, which is not observed for those not selected into the survey. Second, for all individuals selected into the survey, we observe W_2 and A, where the probability of $A = 1$ depends on W_2. Third, selection into the sub-sample (Δ_{sub}) is determined, where the probability of selection depends on W_1 and W_2. Fourth, for those in the sub-sample, we observe one of the two counterfactuals Y_0 or Y_1, corresponding to the treatment actually received.

Figure 4.2: Simulation Set-up. X indicates data present.

			Complete Data				Observed Data			
	Δ_{svy}	Δ_{sub}	**W**	A	$Y^{A=0}$	$Y^{A=1}$	**W**	A	$Y^{A=0}$	$Y^{A=1}$
Population N=100,000	0	0	X	X	X	X				
					X	X				
Survey Sample N=10,000	1	0	X	X	X	X	X	X		
					X	X				
Survey Subsample N=5,000	1	1	X	X	X	X	X	X	X	
					X	X				X

As seen in Figure 4.1, W_2 acts as a confounder. W_1 directly modifies the treatment effect and is related to selection into the survey and sub-sample. Table 4.1 provides summary statistics for the first simulated dataset. The $E(Y_1|A=1) - E(Y_0|A=0)$ values in this table are the naïve estimates of the population average effect, ignoring sample and treatment selection. A consistent estimate of the average effect in the survey sub-sample will require accounting for confounding by W_2. A consistent estimate of the average effect in the population will also require accounting for differential selection by W_1.

Table 4.1: Summary statistics, simulated dataset.

Sample	Variable	Estimate	
sub-sample	W_1 (median, IQR)	0.767	(0.081, 1.509)
	W_2 (median, IQR)	2.341	(1.696, 2.982)
	A ($N^{A=1}$, %)	1943	41.7%
	Y (median, IQR)	6.423	(1.998, 12.108)
	E(Y_1| A=1) - E(Y_0| A=0)	9.714	
survey sample	W_1 (median, IQR)	0.566	(-0.089, 1.254)
	W_2 (median, IQR)	1.984	(1.293, 2.67)
	A ($N^{A=1}$, %)	3658	34.2%
	Y (median, IQR)	4.083	(0.305, 9.305)
	E(Y_1| A=1) - E(Y_0| A=0)	9.175	
population	W_1 (median, IQR)	-0.002	(-0.679, 0.674)
	W_2 (median, IQR)	1.995	(1.322, 2.668)
	A ($N^{A=1}$, %)	34235	34.2%
	Y (median, IQR)	3.837	(0.27, 8.623)
	E(Y_1| A=1) - E(Y_0| A=0)	8.06	

We evaluate how well IPW, TMLE, and DRWLS perform in terms of estimating the population average effect when models are correctly specified and when there is model misspecification (see Tables 4.2 and 4.3). We do not evaluate performance under misspecification of multiple models simultaneously, as this would depend on the particulars of the data generating process and misspecifications. [23] Performance is evaluated by measuring mean percent bias, mean variance, mean-squared error (MSE), and 95% CI coverage across the 1,000 simulations. For each simulation iter-

ation, variance and the 95% CI are estimated from 500 bootstrapped samples. The percentile method is used for the CI.

Table 4.2: Model Misspecification, Scenario 1.

Description	Treatment	Sub-sample Selection	Outcome
Correct Specification	A $\sim$ poly(W_2, 2)	$\Delta_{sub} \sim$ poly(W_1, 2) + W_2	Y $\sim$ A + poly(W_2, 2) + A:W_1
Moderately Misspecified Treatment	A $\sim W_2$	$\Delta_{sub} \sim$ poly(W_1, 2) + W_2	Y $\sim$ A + poly(W_2, 2) + A:W_1
Majorly Misspecified Treatment	A $\sim W_1$	$\Delta_{sub} \sim$ poly(W_1, 2) + W_2	Y $\sim$ A + poly(W_2, 2) + A:W_1
Moderately Misspecified Outcome	A $\sim$ poly(W_2, 2)	$\Delta_{sub} \sim$ poly(W_1, 2) + W_2	Y $\sim$ A + W_2 + A:W_1
Majorly Misspecified (A) Outcome	A $\sim$ poly(W_2, 2)	$\Delta_{sub} \sim$ poly(W_1, 2) + W_2	Y $\sim$ A + A:W_1
Majorly Misspecified (B) Outcome	A $\sim$ poly(W_2, 2)	$\Delta_{sub} \sim$ poly(W_1, 2) + W_2	Y $\sim$ A + poly(W_2, 2) + W_1
Moderately Misspecified Selection	A $\sim$ poly(W_2, 2)	$\Delta_{sub} \sim W_1 + W_2$	Y $\sim$ A + poly(W_2, 2) + A:W_1
Majorly Misspecified Selection	A $\sim$ poly(W_2, 2)	$\Delta_{sub} \sim W_1$	Y $\sim$ A + poly(W_2, 2) + A:W_1

Table 4.3: Model Misspecification, Scenario 2.

Description	Treatment	Sub-sample Selection	Outcome
Correct Specification	A $\sim$ poly(W_2, 2)	$\Delta_{sub} \sim$ poly(W_1, 2) + W_2	Y $\sim$ A + W_2 + A:poly(W_1,2)
Moderately Misspecified Treatment	A $\sim W_2$	$\Delta_{sub} \sim$ poly(W_1, 2) + W_2	Y $\sim$ A + W_2 + A:poly(W_1,2)
Majorly Misspecified Treatment	A $\sim W_1$	$\Delta_{sub} \sim$ poly(W_1, 2) + W_2	Y $\sim$ A + W_2 + A:poly(W_1,2)
Moderately Misspecified Outcome	A $\sim$ poly(W_2, 2)	$\Delta_{sub} \sim$ poly(W_1, 2) + W_2	Y $\sim$ A + W_2 + A:W_1
Majorly Misspecified (A) Outcome	A $\sim$ poly(W_2, 2)	$\Delta_{sub} \sim$ poly(W_1, 2) + W_2	Y $\sim$ A + W_2 + poly(W_1, 2)
Majorly Misspecified (B) Outcome	A $\sim$ poly(W_2, 2)	$\Delta_{sub} \sim$ poly(W_1, 2) + W_2	Y $\sim$ A + W_2 + W_1
Moderately Misspecified Selection	A $\sim$ poly(W_2, 2)	$\Delta_{sub} \sim W_1 + W_2$	Y $\sim$ A + W_2 + A:poly(W_1,2)
Majorly Misspecified Selection	A $\sim$ poly(W_2, 2)	$\Delta_{sub} \sim W_1$	Y $\sim$ A + W_2 + A:poly(W_1,2)

4.4.2 Results

Table 4.4 provides a summary of method performance under correct model specification and under model misspecification. The DRWLS and TMLE estimators perform similarly and outperform the IPW estimator in nearly all cases. When all models are correctly specified, DRWLS and TMLE have lower MSE than IPW, lower vari-

ance, and greater 95% CI coverage (see Table 4.5). The advantages of DRWLS and TMLE over IPW become especially pronounced under misspecification of the treatment model. This result is expected, because IPW relies exclusively on the inverse probability of treatment weights to account for non-random treatment assignment. In contrast, the DRWLS and TMLE estimators are consistent under misspecification of the treatment model if the outcome model is correctly specified.

Several authors have warned that IPW and double robust estimators, like TMLE, are sensitive in scenarios of practical positivity violations—i.e., when subsets of the sample have very large weights. [9, 22–24, 28] The maximum estimated selection-treatment weight in our simulations is 2,000, which—while not as extreme as those in Kang and Schafer (2007) [9]—is regarded as a practical positivity scenario. [21] Table 4.6, below, gives the range of true and estimated conditional probabilities of selection and treatment in the first simulated dataset. In addition, IPW estimators may be problematic if those not in the sample/untreated have weights that are many times greater than those in the sample/treated, as this also leads to more reliance on model extrapolation. In our simulated dataset, the minimum weight among those treated and selected into the sub-sample was 3.2 times that of those untreated and not selected into the sub-sample. The maximum weight among those treated and selected into the sub-sample was 1.7 times that of those untreated and not selected.

Table 4.4: Method performance under correct specification and misspecification. ▲= good,◆=fair, ▼=poor

	IPW				DRWLS				TMLE			
	% bias	var	cov	MSE	%bias	var	cov	MSE	% bias	var	cov	MSE
Scenario 1												
True Model	▲	◆	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Mod MisTx	▼	▼	▲	◆	▲	▲	▲	▲	▲	▲	▲	▲
Maj MisTx	▼	▲	▼	▼	▲	▲	▲	▲	▲	▲	▲	▲
Mod MisOut	▲	◆	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Maj MisOut	▲	◆	▲	▲	▲	◆	▲	◆	▲	◆	▲	◆
Mod MisSel	▼	▼	▲	▼	▲	▲	▲	▲	▲	▲	▲	▲
Maj MisSel	▼	▼	▲	▼	▲	▲	▲	▲	▲	▲	▲	▲
Scenario 2												
True Model	▲	◆	▲	◆	▲	▲	▲	▲	▲	▲	▲	▲
Mod MisTx	▼	◆	▲	◆	▲	▲	▲	▲	◆	◆	▲	◆
Maj MisTx	▼	▲	▼	▼	▲	▲	▲	▲	▲	▲	▲	▲
Mod MisOut	▲	◆	▲	◆	▲	▲	▲	▲	▲	▲	▲	▲
Maj MisOut	▲	◆	▲	◆	▲	▲	▲	▲	▲	▲	▲	◆
Mod MisSel	▲	◆	▲	◆	▲	▲	▲	▲	▲	▲	▲	▲
Maj MisSel	▲	▲	▲	◆	▲	▲	▲	▲	▲	▲	▲	▲

Table 4.5: Method performance under correct specification and misspecification. Mean % bias, mean variance (Var), 95% CI coverage (Cov), and mean-squared error (MSE) across the 1,000 simulations.

	IPW				DRWLS				TMLE			
	% Bias	Var	Cov	MSE	% Bias	Var	Cov	MSE	% Bias	Var	Vov	MSE
Scenario 1												
TrueMod	0.3	0.148	91.5	0.156	0.4	0.014	95.8	0.014	0.4	0.015	95.8	0.014
ModMisTx	-4.0	0.260	91.0	0.343	0.3	0.015	95.2	0.016	0.0	0.017	95.9	0.016
MajMisTx	168.6	0.038	0.0	25.644	0.3	0.010	95.8	0.009	0.2	0.009	96.5	0.008
ModMisOut1	0.0	0.000	0.0	0.000	-0.3	0.028	93.3	0.033	0.2	0.032	92.9	0.035
MajMisOut2	0.0	0.000	0.0	0.000	0.3	0.115	91.3	0.121	-0.3	0.073	94.6	0.074
MajMisOut3	0.0	0.000	0.0	0.000	0.6	0.015	93.6	0.016	0.1	0.023	93.3	0.027
ModMisSel	-11.2	0.330	92.2	0.551	0.3	0.016	94.6	0.017	0.3	0.016	95.3	0.016
MajMisSel	-9.7	0.330	93.0	0.476	0.5	0.012	95.4	0.012	0.4	0.011	94.8	0.011
Scenario 2												
TrueMod	0.8	0.049	91.6	0.049	0.4	0.014	95.3	0.013	0.7	0.035	94.4	0.032
ModMisTx	-6.4	0.153	93.0	0.248	0.1	0.017	95.3	0.019	-2.5	0.053	93.0	0.061
MajMisTx	62.7	0.013	0.0	3.555	0.4	0.010	96.0	0.009	0.1	0.009	96.1	0.007
ModMisOut1	0.0	0.000	0.0	0.000	0.2	0.018	95.5	0.017	0.7	0.035	94.5	0.032
MajMisOut2	0.0	0.000	0.0	0.000	0.4	0.015	95.2	0.014	0.2	0.039	91.0	0.056
MajMisOut3	0.0	0.000	0.0	0.000	0.4	0.017	94.6	0.016	0.2	0.046	91.0	0.065
ModMisSel	-0.5	0.078	92.7	0.084	0.3	0.016	95.0	0.015	0.1	0.015	95.5	0.014
MajMisSel	0.6	0.043	93.3	0.043	0.5	0.012	95.8	0.011	0.2	0.010	95.7	0.009

Table 4.6: Range of true and estimated conditional selection and treatment probabilities in the first simulated dataset.

	True	Estimated
$\mathrm{I}(\Delta_{svy}=1)\mathrm{P}(\Delta_{svy}=1\|W_1)$	0.0106-0.7107	NA
$\mathrm{I}(A=1)\mathrm{P}(A=1\|\Delta_{svy}=1,W_2)$	0.0635-0.9959	0.0584-0.9949
$\mathrm{I}(A=0)\mathrm{P}(A=0\|\Delta_{svy}=1,W_2)$	0.0308-0.9385	0.0344-0.9467
$\mathrm{I}(\Delta_{sub}=1)\mathrm{P}(\Delta_{sub}=1\|\Delta_{svy}=1,W_1,W_2\ )$	0.0628-0.9885	0.0608-0.9869
$\mathrm{I}(\Delta_{sub}=1,A=1)\mathrm{P}(\Delta_{sub}=1,A=1\|\Delta_{svy}=1,W_1,W_2\ )$	0.0095-0.9279	0.0091-0.9241
$\mathrm{I}(\Delta_{sub}=1,A=0)\mathrm{P}(\Delta_{sub}=1,A=0\|\Delta_{svy}=1,W_1,W_2\ )$	0.0039-0.8482	0.0034-0.8475
$\mathrm{I}(\Delta_{sub}=1,A=1,\Delta_{svy}=1)$		
$\times\ \mathrm{P}(\Delta_{sub}=1,A=1,\Delta_{svy}=1\|W_1,W_2\ )$	0.0005-0.4076	0.0005-0.4052
$\mathrm{I}(\Delta_{sub}=1,A=0,\Delta_{svy}=1)$		
$\times\ \mathrm{P}(\Delta_{sub}=1,A=0,\Delta_{svy}=1\|W_1,W_2\ )$	0.0029-0.5665	0.0027-0.5619

We examined the extent to which there is a penalty for unnecessary adjustment for non-random treatment assignment or sample selection. We considered two scenarios for each of the treatment and sub-sample selection models (see Table 4.7). In this limited simulation, there are no noticeable penalties for over-adjusting. Table 4.8 shows the results from the more extreme second scenario. Results from the first scenario were similar.

Table 4.7: Model Misspecification, Overadjustment.

Description	Treatment	Sub-sample Selection	Outcome
Moderate Over-adjustment, Treatment			
True	A $\sim W_2$	$\Delta_{sub} \sim$ poly(W_1, 2) + W_2	Y $\sim$ A + poly(W_2, 2) + A:W_1
Misspecified	A $\sim$ poly(W_2, 2)	$\Delta_{sub} \sim$ poly(W_1, 2) + W_2	Y $\sim$ A + poly(W_2, 2) + A:W_1
Major Over-adjustment, Treatment			
True	random	$\Delta_{sub} \sim$ poly(W_1, 2) + W_2	Y $\sim$ A + poly(W_2, 2) + A:W_1
Misspecified	A $\sim$ poly(W_2, 2)	$\Delta_{sub} \sim$ poly(W_1, 2) + W_2	Y $\sim$ A + poly(W_2, 2) + A:W_1
Moderate Over-adjustment, Selection			
True	A $\sim$ poly(W_2, 2)	$\Delta_{sub} \sim W_1 + W_2$	Y $\sim$ A + poly(W_2, 2) + A:W_1
Misspecified	A $\sim$ poly(W_2, 2)	$\Delta_{sub} \sim$ poly(W_1, 2) + W_2	Y $\sim$ A + poly(W_2, 2) + A:W_1
Major Over-adjustment, Selection			
True	A $\sim$ poly(W_2, 2)	$\Delta_{sub} \sim$ random	Y $\sim$ A + poly(W_2, 2) + A:W_1
Misspecified	A $\sim$ poly(W_2, 2)	$\Delta_{sub} \sim$ poly(W_1, 2) + W_2	Y $\sim$ A + poly(W_2, 2) + A:W_1

Table 4.8: Results under misspecification of the treatment and selection models: adjustment when the treatment and selection mechanisms are completely random. Mean % bias, mean variance (Var), 95% CI coverage (Cov), and mean-squared error (MSE) across the 1,000 simulations.

	Correct Specification			Overadjustment		
	IPW	DRWLS	TMLE	IPW	DRWLS	TMLE
Treatment						
% Bias	0.3	0.0	0.0	-0.2	0.0	0.0
Var	0.043	0.009	0.009	0.032	0.009	0.009
Cov	96.3	95.5	95.8	94.8	95.5	95.8
MSE	0.039	0.008	0.008	0.031	0.008	0.008
Selection						
% Bias	0.3	0.3	0.3	0.2	0.3	0.3
Var	0.121	0.008	0.009	0.108	0.008	0.009
Cov	92.3	94.7	95.0	91.2	95.0	95.1
MSE	0.132	0.008	0.008	0.120	0.008	0.008

4.5 Case Study

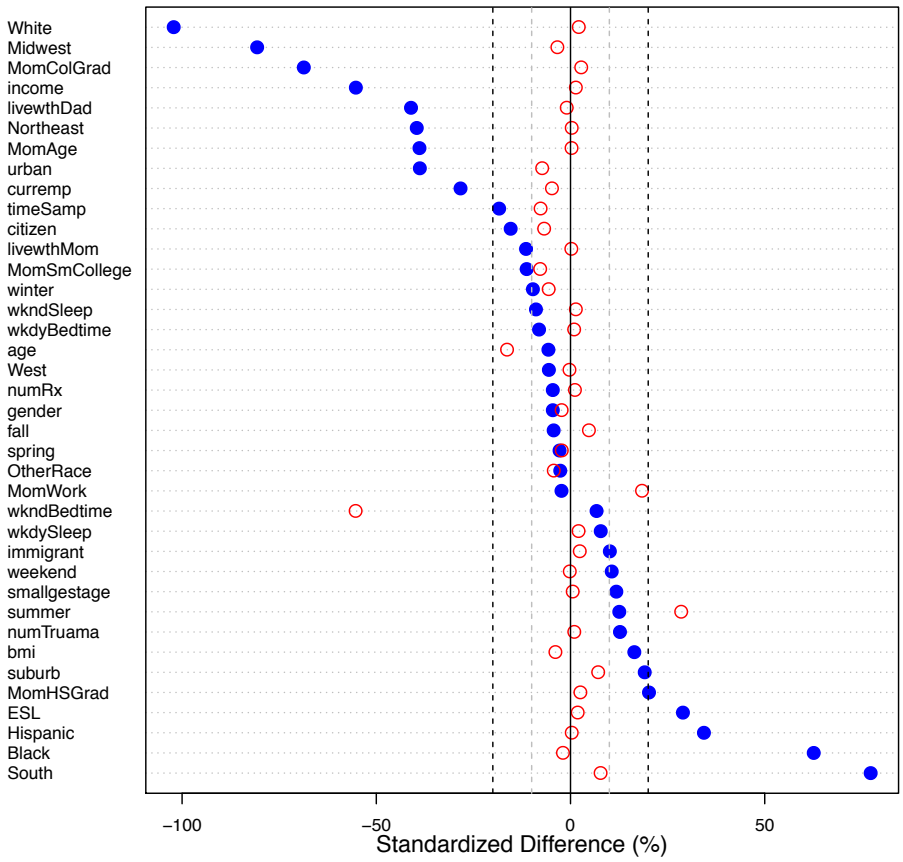
4.5.1 Overview and set-up

We now apply the estimators evaluated in the above simulation to the motivating case study. As stated in the Introduction, we want to generalize the effect of disadvantaged neighborhood residence on cortisol slope to the population of U.S. adolescents. The NCS-A has been described previously. [10–12,16] Each adolescent was interviewed in her/his home by a professional interviewer. Interviews lasted an average of 2 1/2 hours. Neighborhood disadvantage was measured using an established scale [6] that has been used previously with NCS-A residence data geocoded to Census tracts. [27] Salivary cortisol samples, a hormone involved in the hypothalamic-pituitary-adrenal axis, [15] were taken immediately before and after the survey interview. Cortisol samples were analyzed for a sub-sample of 2,490 participants because of budget lim-

itations. Exposure and covariate data were available for all participants. Analysis of the relationship between neighborhood disadvantage and cortisol slope among the sub-sample of participants with cortisol data has been previously reported. [27] Informed assent and consent were obtained from each adolescent and his/her parent or guardian. The Human Subjects Committees of Harvard Medical School and the University of Michigan approved recruitment and consent procedures.

Figure 4.3 depicts the extent to which 1) NCS-A participants with cortisol measurements compare to participants without across possible confounding variables (red, open dots), and 2) NCS-A participants with the exposure of interest (residence in a disadvantaged neighborhood) compare to participants without (blue, closed dots). Those with and without cortisol measurements look similar with the exceptions of age, average bedtime on the weekends, maternal work, and the interview taking place in the summer. In contrast, participants living in disadvantaged neighborhoods differ from those living in non-disadvantaged neighborhoods in terms of expected demographic variables like race, income, and maternal education. We do not see any variables for which the red and blue dots are both extreme.

Figure 4.3: Covariate balance. Solid points represent the standardized mean differences between the disadvantaged neighborhood group and non-disadvantaged neighborhood group. Open points represent the standardized mean differences between those with cortisol measurement and those without. The standardized mean difference is the difference in means between the two groups standardized by the standard deviation in the first group.



The estimated weights for these example data are shown in Table 4.9. Positivity violations can be a substantial issue in observational studies. [5, 17] The survey weights in our case study are large, reflecting the NCS-A survey design. As discussed, double robust estimators may rely on model extrapolation and the variance of IPW estimators can become large if weights for the participants who were treated and

selected into the sub-sample are many times larger than weights for the participants who were untreated and not selected. In this case study, the minimum weight among those treated and selected into the sub-sample was 1.7 times that of those untreated and not selected when survey weights were ignored, and 1.3 times that when survey weights were used. The maximum weight among those treated and selected into the sub-sample was 1.6 times that of those untreated and not selected when survey weights were ignored and 1.3 times that when survey weights were used.

Table 4.9: Range of estimated selection and treatment probabilities conditional on covariates.

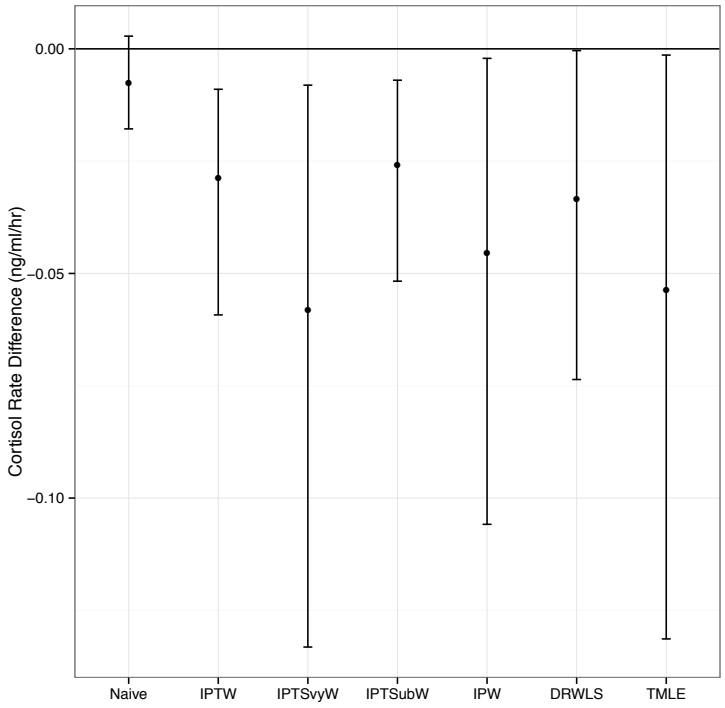
	Estimated
$\mathrm{I}(\Delta_{svy}=1)\mathrm{P}(\Delta_{svy}=1\|\mathbf{W})$	4×10^{-5}-0.0122
$\mathrm{I}(A=1)\mathrm{P}(A=1\|\Delta_{svy}=1,\mathbf{W})$	0.0422-0.9784
$\mathrm{I}(A=0)\mathrm{P}(A=0\|\Delta_{svy}=1,\mathbf{W})$	0.0337-0.9851
$\mathrm{I}(\Delta_{sub}=1)\mathrm{P}(\Delta_{sub}=1\|A=1,\Delta_{svy}=1,\mathbf{W})$	0.0420-0.7379
$\mathrm{I}(\Delta_{sub}=1)\mathrm{P}(\Delta_{sub}=1\|A=0,\Delta_{svy}=1,\mathbf{W})$	0.0433-0.6046
$\mathrm{I}(\Delta_{sub}=1,A=1)\mathrm{P}(\Delta_{sub}=1,A=1\|\Delta_{svy}=1,\mathbf{W})$	0.0062-0.5968
$\mathrm{I}(\Delta_{sub}=1,A=0)\mathrm{P}(\Delta_{sub}=1,A=0\|\Delta_{svy}=1,\mathbf{W})$	0.0063-0.4873
$\mathrm{I}(\Delta_{sub}=1,A=1,\Delta_{svy}=1)\mathrm{P}(\Delta_{sub}=1,A=1,\Delta_{svy}=1\|\mathbf{W})$	1.3×10^{-6}-0.0051
$\mathrm{I}(\Delta_{sub}=1,A=0,\Delta_{svy}=1)\mathrm{P}(\Delta_{sub}=1,A=0,\Delta_{svy}=1\|\mathbf{W})$	1.5×10^{-6}-0.0040

4.5.2 Results

Figure 4.4 plots the estimates and 95% CIs for the expected effect of living in a disadvantaged neighborhood on cortisol slope using different bias-correction methods. The 95% CIs are calculated by the percentile method using 1,000 bootstrapped

samples.

Figure 4.4: Illustrative example: marginal mean effect estimates and 95% confidence intervals



We first present simpler methods that adjust for none or only some of the potential biases that exist (in comparison to the IPW, TMLE, and DRWLS methods described above, which can simultaneously handle non-random survey selection, sub-sample selection, and treatment assignment). Under the naÃŕve approach (no bias correction), the point estimate is negative and the confidence interval is narrow and crosses zero. The IPTW estimator adjusts for non-random assignment of the treatment but does not address sample selection. If the treatment model is correctly specified, the IPTW estimate will be consistent for the average effect in the NCS-A sub-sample but may be biased for the average effect in the population. The IPTSvyW estimator adjusts for non-random assignment of the treatment and non-random selection into the survey,

but it does not adjust for non-random selection into the sub-sample. The IPTSubW estimator adjusts for non-random assignment of the treatment and non-random selection into the survey sub-sample. If the two models are correctly specified, IPTSubW will be a consistent estimate of the average effect in the NCS-A survey sample but may be biased for the average effect in the target population.

One-dimensional summaries (Figure 4.3) may lead us to believe that it is not important to adjust for selection into the sub-sample. The IPW estimate that adjusts for this selection (IPW) is more than 22% larger than that which does not (IPTSvyW). This case study was chosen because it was the original motivation for this paperâĂŤit does not demonstrate inferential differences when sub-sample selection is or is not adjusted for. Nonetheless, it is possible that even small differences in a plot like Figure 3 could result in substantially biased population average effect estimates.

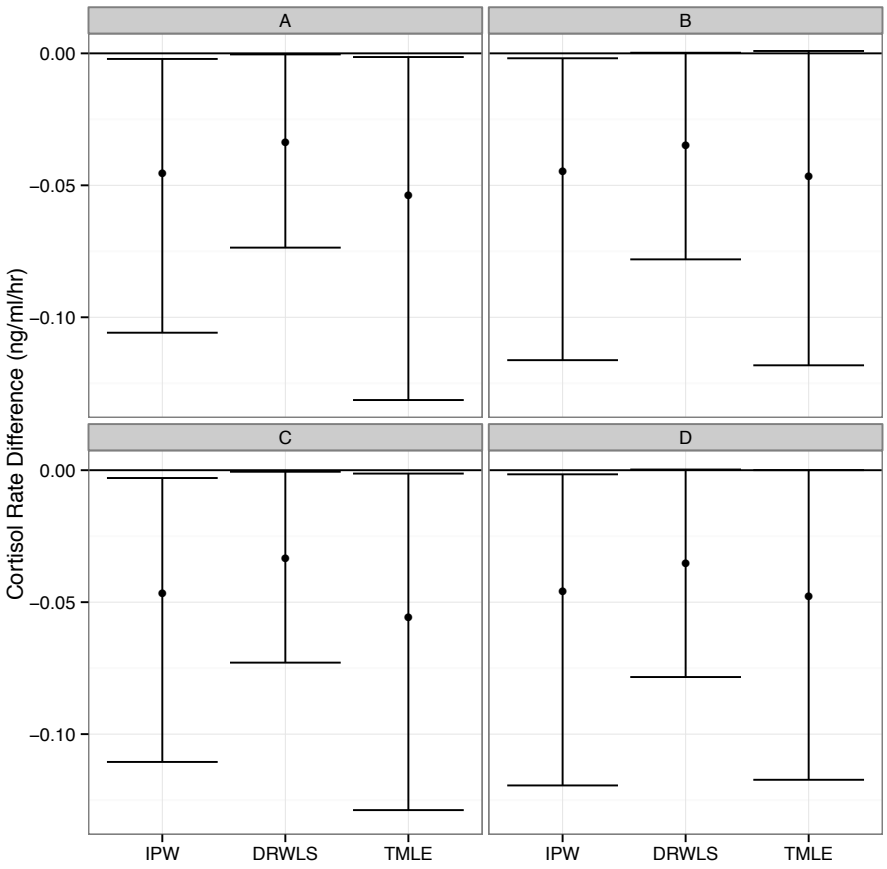
The IPW, TMLE, and DRWLS estimators adjust for non-random assignment of the treatment, non-random selection into the survey sample, and non-random selection into the sub-sample. TMLE and DRWLS have the additional advantage that only two of the three treatment, sub-sample selection, and outcome regression models need to be correctly specified to consistently estimate the population average effect. In our simulation, both TMLE and DRWLS had smaller variance than IPW. In this case study, however, only DRWLS has smaller variance than IPW, which results in a 29% narrower confidence interval. The discrepancy between the performance of TMLE in the simulation and in the case study likely stems from more unstable weights and practical positivity violations in the case study data. This highlights the need for caution in such scenarios.

Based on our simulation study, we would anticipate the DRWLS and TMLE estimates to be consistent if only one of our models were incorrectly specified. Using

DRWLS, we conclude that the cortisol rate difference comparing U.S. adolescents in disadvantaged versus non-disadvantaged neighborhoods likely falls between -7.36 and -0.04 $\times 10^{-2}$ ng/mL/hour.

Just as we assess whether there are penalties for unnecessarily adjusting for non-random treatment and non-random sub-sample selection in the simulation study (see Table 4.8), we compare case study results using more parsimonious and less parsimonious models. We chose the models using Akaike information criterion (AIC) in a stepwise algorithm. Using the AIC may result in models that have better predictive ability but are less parsimonious. In the simulation, we paid no noticeable penalty for over-adjusting or unnecessarily adjusting for non-random treatment and non-random selection. It is possible, though, that we may pay an efficiency penalty in the more complicated real-world case study. To examine this, we compare A) the IPW, TMLE, and DRWLS estimators that use the full, AIC-optimized treatment and selection models to B) those that use the full, AIC-optimized treatment model and a more parsimonious Bayesian information criterion (BIC)-optimized selection model, C) those that use the BIC-optimized treatment model and the AIC-optimized selection model, and D) those that use the BIC-optimized selection and treatment models (see Figure 4.5). When the more parsimonious selection model is used in panels A and D, the IPW confidence intervals slightly widen whereas the TMLE confidence intervals slightly narrow, resulting in similar confidence interval widths comparing IPW to TMLE. We see little difference between panels A-D in terms of the DRWLS estimate and 95% CIs.

Figure 4.5: Illustrative example: marginal mean effect estimates and 95% confidence intervals under different levels of parsimony in model specification.



4.6 Discussion

We evaluated estimators of the population average effect in the presence of treatment effect heterogeneity, non-random treatment assignment, and a two-stage selection process. Using a simulation study we found that a DRWLS estimator and a TMLE estimator have lower MSE, variance, and percent bias than the IPW estimator. This was true when all models were correctly specified and under moderate and major misspecification of one of the treatment, selection, or outcome models. We

derive the efficient influence function and present a TMLE estimator incorporating survey sampling weights, which can be easily implemented using the available tmle package in R.

We agree with others [4, 23, 28, 30] that estimating an average effect standardized to a population of interest is a practical goal. It can aid in the interpretability and applicability of a studyâĂŹs conclusions, provided one recognizes the assumptions and limitations involved. First, a population average effect will not provide information about treatment effect heterogeneity. Second, estimation can be difficult in the presence of positivity violations. In cases where the weights are highly variable, Robins et al. (2007) recommend a sensitivity analysis varying specification of the models. [23] In addition, there exist methods to identify possible instances of biases due to positivity violations. [20, 34] Non-parametric methods of model specification may improve robustness to model misspecification. [3]

Our demonstration of the poor performance of IPW is not new. IPW estimators have well-known efficiency problems and can be biased due to structural or practical positivity violations. [23] Much has been published on this in the biostatistics literature (e.g., [23, 24]), but IPW continues to be widely used by epidemiologistsâĂŤperhaps because it is straightforward to implement in standard statistical software. We hope by demonstrating the similarly straightforward implementation of DRWLS and TMLE coupled with their superior performance over IPW, use of these estimators may gain popularity.

We evaluated the robustness of our simulation results in a series of sensitivity analyses. First, we truncated the combined treatment and selection weights at the 99th percentile, as this may lessen both bias and variance due to extreme weights, though it may also increase bias due to misspecification. [22] We simplified from a

two-stage to a one-stage selection process. We also tested robustness to weakening and strengthening the relationship between W_1 and selection. Method performance under correct model specification did not change in any of these cases. We also modified the data generating mechanism in the simulation so that both covariates were associated with probability of treatment and probability of selection. IPW performance improved under this scenario but was still less efficient than TMLE or DRWLS. Finally, we repeated simulations under no effect heterogeneity and binary effect heterogeneity (W_1 Bernoulli(1/2)). Performance of TMLE and DRWLS did not change much under these scenarios. However, the MSE of IPW improved in the case of binary effect heterogeneity and improved even further in the case of no effect heterogeneity, possibly due to fewer practical positivity violations. [18]

In this paper's simulation and example, we considered a scenario where the full set of covariates was measured in the larger survey sample and selection into the sub-sample only affected missingness of the outcome variable. One could also conceive of scenarios where the monotone missing data patterns extends to some subset of covariates. For example, let the full set of covariates be represented by $\mathbf{W} = W_1, W_2$. Participants in the sub-sample have the full set of covariates W_1, W_2, but participants in the survey sample have only the subset W_1. In this case, the observed data would be $O = (W_1,\ \Delta_{svy} = 1,\ A,\ \Delta_{sub},\ \Delta_{sub}W_2,\ \Delta_{sub}Y)$. We explain how In order for the exchangeability assumption to hold, Δ_{svy} and Δ_{sub} must be random conditional on W_1, which implies $P(\Delta_{svy} = 1|W_1) = P(\Delta_{svy} = 1|W_1, W_2)$ and $P(\Delta_{sub} = 1|\Delta_{svy} = 1, W_1) = P(\Delta_{svy} = 1|\Delta_{svy} = 1, W_1, W_2)$. Some additional assumptions and modifications would be required for each of the three estimators evaluated to maintain their unbiasedness properties, as described below.

For the IPW estimator, if W_1 and W_2 are both confounders of the treatment

effect, then the conditional probabilities with which we construct the combined inverse probability of treatment and selection weights would need to change to:

$$\begin{aligned} w^{A=1,\Delta_{svy}=1,\Delta_{sub}=1} &= \frac{1}{P(\Delta_{svy}=1|W_1)} \times \frac{1}{P(\Delta_{sub}=1|\Delta_{svy}=1,W_1)} \\ &\quad \times \frac{1}{P(A=1)|\Delta_{svy}=1,\Delta_{sub}=1,W_1,W_2)} \\ &= \frac{1}{P(A=1,\Delta_{sub}=1,\Delta_{svy}=1|\mathbf{W})} \end{aligned}$$

The IPW estimate would be unbiased if the treatment and selection models were correctly specified. For the TMLE and DRWLS estimators, if W_1 and W_2 are both confounders of the treatment effect, then these estimators would no longer be robust to misspecification of the selection model. This is because the G-computation step could only standardize to the sub-sample, meaning that the predicted individual treatment effects would need to be weighted by the combined survey weights and sub-sample weights to compute the marginal mean effect estimate for the population. TMLE and DRWLS would maintain unbiasedness under correct model specification and either treatment or outcome model misspecification.

Our simulation study has some limitations. First, simulations can only give a rough approximation of the sampling distribution of the estimators. [23] Second, there are arguably nearly a dozen or more estimators that could have been assessed and compared. [13, 14, 28] We chose to focus on a smaller set of estimators that are particularly straightforward to implement. It is an area for future work to develop easy-to-use software packages implementing these estimators, as has been done in the case of TMLE. [7] Third, the approach shown in this paper is not a fully design-based survey analysis. For example, we ignore survey sampling strata in our bootstrapping

procedure. This is another area for future work.

In conclusion, we compared estimators of an average effect standardized to a target population in the presence of non-random treatment assignment, a two-stage selection process, and treatment effect heterogeneity (linear and non-linear). This scenario can apply to generalizing results from a survey sub-sample to a specified target population. [4, 30] We demonstrated that DRWLS and TMLE estimators outperform an IPW estimator in terms of percent bias, variance, and MSE, even under misspecification of one of the treatment, selection, or outcome models. Moreover, they are similarly easy to implement. Lastly, we demonstrated how DRWLS and TMLE estimators can be applied to everyday research questions, providing an attractive alternative to IPW for applied epidemiologic researchers.

Chapter 5

Discussion

5.1 Goal

Previous research into the relationship between neighborhood disadvantage and adolescent stress and depression and anxiety has resulted in inconsistent evidence. In this dissertation, we hypothesize that several factors could contribute to these inconsistencies, including (1) nontransportability of findings due to effect heterogeneity and nonrandom sample selection, (2) failure to fully address confounding by nonrandom neighborhood assignment, and (3) extrapolation induced by nonpositivity. This dissertation aims to address these gaps using causal inference methods. Specifically, this dissertation examines effect heterogeneity in the relationships of neighborhood disadvantage and (1) cortisol levels and (2) prevalent depression/anxiety in adolescents and suggests methods to address such heterogeneity. In addition, we use propensity score methods and sensitivity analyses to address both observed and unobserved confounding by nonrandom neighborhood assignment and positivity violations.

5.2 Findings

5.2.1 Effect Modification

We demonstrate that neighborhood disadvantage is associated with greater prevalence of depression/anxiety if the neighborhood is in an urban center, but not if the neighborhood is in a suburban or rural area. This is congruent with the pattern of inconsistent results seen in the literature. Articles reporting no association tended to sample study participants from the suburbs or rural areas,[e.g., [8, 30]] but those reporting an association tended to sample the participants from urban areas.[e.g., [18, 20, 45]] Thus, it is possible that considering urbanicity as an effect modifier may offer a partial explanation of previous inconsistent findings. However, we consider alternative explanations in Section 5.3.1.1.

Effect modifiers may be integral components of the theoretical mechanism underlying an association. For example, the theory of how neighborhood disadvantage is embodied to affect mental health is likely different depending on whether the process occurs in an urban center versus a rural area. Living in a disadvantaged neighborhood in an urban area may include more exposure to neighborhood violence and noise [e.g., [7, 31]] and more economic segregation in neighborhoods, schools, and workplaces. [41] In such cases, it is important to identify effect modifiers and estimate subgroup effects for at least two reasons. First, the Stable Unit Treatment Value Assumption (SUTVA) of causal inference assumes that there is one version of the treatment/exposure for everyone. [28] Thus, examining the effect of a treatment that is as homogenous as possible is likely useful in determining what is it about neighborhood that increases risk of anxiety and depression. Second, estimating sub-

group effects where the subgroups are defined based on the value of an effect modifier is important for targeting interventions to individuals who may benefit most. For instance, in our example of the relationship between neighborhood disadvantage and prevalent depression/anxiety, adolescents in disadvantaged urban areas were found to be particularly vulnerable to anxiety and depression. Therefore, targeting resources to this subpopulation may be appropriate.

When effect modifiers are not integral to the theoretical mechanism, they may still be important to consider because of their impact on the generalizability of findings. For example, in Aim 2 we find that living in a disadvantage neighborhood is associated with higher cortisol levels prior to a novel interview situation and a steeper rate of decline in cortisol levels over the course of the interview. However, cortisol data were only available for a nonrandom subset of the nationally representative sample, so we could not be sure whether our results would apply to U.S. adolescents in general. Hernan and Robins call the ability to apply causal effects estimated in one population to other populations "transportability". [14] If populations differ in the distribution of effect modifiers, and if these effect modifiers are not accounted for in the analysis (e.g., as interactions with treatment in a regression), and/or if a marginal effect is estimated, then effect estimates will differ between populations and such differences may be large enough to cause changes in inference. Nontransportability due to nonrandom sample selection is discussed less frequently than confounding due to nonrandom treatment/exposure assignment. In epidemiologic literature, it is typically addressed by inverse probability weighting, if it is addressed at all. [4] However, these weighting methods have been shown to be problematic—particularly in terms of efficiency, but even in terms of bias. [25] We hypothesize that inverse probability weighting (IPW) is still frequently used by epidemiologists because of a lack of dissemination about

the problems of weights and because weights are easy to implement.

In this dissertation, we demonstrate that two easy-to-implement versions of targeted maximum likelihood estimation (TMLE) can be used to address bias due to nonrandom sampling (and thus address nontransportability issues) as well as bias due to nonrandom treatment/exposure assignment. These two versions of TMLE include a simple estimator suggested by Marshall Joffe that combines weighted regression with G-computation and an estimator that uses a vector of inverse probability weights to fluctuate around an initial conditional expected outcome. We demonstrate via simulation that in the scenario of nonrandom treatment assignment, a nonrandom two-stage selection mechanism, and treatment effect heterogeneity (both linear and nonlinear), these TMLE implementations outperform IPW in terms of mean squared error and bias. The advantages of TMLE over IPW hold under no model misspecification as well as in the presence of misspecification of any one of the treatment, selection, or outcome models. To facilitate adoption of this method by epidemiologists and other applied researchers, we provide a tutorial of how both TMLE implementations may be implemented, including code.

5.2.2 Confounding and Positivity

In addition to effect heterogeneity, this dissertation addresses confounding by neighborhood residence and positivity violations in estimating the associations between neighborhood disadvantage and anxiety/depression and between neighborhood disadvantage and cortisol. Propensity score methods—particularly when applied in conjunction with usual outcome analytic methods like regression—offer several advantages over regression alone, as discussed in Section 1.5.1. We applied propensity score

subclassification and coupled it with regression adjustment to make the assumptions of no confounding and positivity more tenable in the association between neighborhood disadvantage and adolescent depression/anxiety in Aim 1. We also combined propensity score methods with multiple imputation and design-based survey analysis in this Aim's analyses. Approaches for combining these methods have not been well studied, and this is an area of interest for future methodological work. For example, we know of few published examples that have combined propensity scores and a design-based survey analysis. [5, 46] We largely followed Zanutto's approach, but employed the Mantel-Haenszel method in averaging over propensity score subclasses, which resulted in smaller standard errors. Future work should explore whether this method would result in efficiency gains more generally.

We also demonstrate how multiple matching methods can be combined and coupled with regression to provide a way to reduce variability in a large, nonlaboratory-based study involving biomarkers in Aim 2. This is relevant to neighborhood research, because if neighborhood residence influences risk of depression and anxiety, its influence may operate through the stress-response system: living in a disadvantaged neighborhood may increase stress levels, which place residents at differentially greater risk of poor mental health. Neighborhood studies may encompass large geographic areas to include ample neighborhood variability in the sample. Study participants drawn from a large geographic area present a challenge to studying stress, however, because it may be logistically and financially prohibitive to transport participants to a central location so that their stress biomarkers can be measured in a strictly controlled laboratory setting. Studies of adults have addressed this challenge by having participants take multiple cortisol samples at certain times per day on multiple days to separate within- versus between-person variability in cortisol's diurnal rhythm. [11] However,

this collection scheme has been shown to be infeasible for adolescents. [12] Moreover, a standardized stress test was deemed to be unethical to administer to adolescents in the NCS-A (and presumably in other studies involving children). Consequently, the NCS-A measured cortisol at the time of the interview in the adolescent's home. To take advantage of these data, including the major strength of racial/ethnic and geographic diversity and large sample size, we needed to take steps to reduce variability in the cortisol data. Key drivers of cortisol variability include time of day, weekday versus weekend, and possibly race/ethnicity. To achieve strict balance on these variables between the two exposure groups, we matched individuals living in disadvantaged neighborhoods with those in nondisadvantaged neighborhoods who had the same values of weekend/weekday and race/ethnicity (exact matching), who had their cortisol measured at approximately the same time of day (caliper matching), and who looked similar on a lengthy vector of demographic characteristics and environmental and behavioral characteristics that may also affect cortisol levels (e.g., season, bedtime) (propensity score matching). Prior to matching, we applied strict exclusion criteria to exclude individuals whose stress response system may be influenced by hormones or drugs and so may not be at risk for being influenced by neighborhood sources of stress. This variability-reducing strategy may provide an initial way forward for the practical researcher wanting to make use of cortisol measurements in large, epidemiologic studies.

Because Aims 1 and 2 are observational studies, we cannot be sure that the associations are not actually artifacts of unobserved confounding. Consequently, for each Aim, we conducted sensitivity analyses to estimate how much unobserved confounding would be necessary to change our inferences. There have been many articles on how to estimate the impact of unobserved confounding, but most are specific to certain,

simple scenarios. [26] We used the more general equations proven in VanderWeele and Arah. [39] We found that the association between neighborhood disadvantage and adolescent depression/anxiety in an urban area was not very sensitive to an unobserved confounder. (Setting the conditional probability of a given unobserved confounder to 80% in disadvantaged neighborhoods conditional on a vector of measured covariates, $\mathbf{X}$, the unobserved confounder would need to be twice as prevalent in disadvantaged versus nondisadvantaged neighborhoods (conditional on $\mathbf{X}$) and be associated with 1.4 times greater odds of prevalent depression/anxiety (conditional on $\mathbf{X}$) to change our inference.) However, the association between neighborhood disadvantage and cortisol slope over the course of the NCS-A interview was moderately sensitive to an unobserved confounder. (Setting the conditional probability of a given unobserved confounder to be 20% greater in a disadvantaged versus nondisadvantaged neighborhoods (conditional on $\mathbf{X}$), the unobserved confounder would have to change the conditional average effect of neighborhood on cortisol slope by 62% to change our inference in Adjusted Model 1.)

5.3 Limitations

As discussed in the above paragraph, because this dissertation estimated associations with a nonrandomized exposure, one limitation is that results may be biased due to unobserved confounding. Other significant limitations—both practical and philosophical—stem from our use of neighborhood disadvantage as the exposure of interest.

5.3.1 Error in measuring neighborhood disadvantage

5.3.1.1 Practical limitations

First, neighborhood disadvantage is measured with error. The summary measure of neighborhood disadvantage comprises a set of indicators measuring different components of neighborhood disadvantage, like income, assets, housing value, education, and employment. [27] While a summary measure such as this one is more complete and should contain less random measurement error than any one indicator alone, it is unlikely that the latent construct of neighborhood disadvantage is fully captured by this set of indicators. Furthermore, we use Census tracts as a proxy for neighborhood; neighborhood boundaries identified by residents will likely not overlap completely with Census tract boundaries. In addition, the quality of Census tracts as a proxy for neighborhood may differ by level of urbanicity or neighborhood-level effect modifer. We discuss this limitation further in Section 5.3.1.1. Nevertheless, Census tracts allow neighborhood measures to be "compared over time and across regions," [19] and are better than zip codes at detecting differences in socioeconomic gradation across areas. [19]

Second, in estimating each association we impose a cutpoint and define neighborhoods in the first tertile of neighborhood disadvantage scores as disadvantaged and those in the upper two tertiles as nondisadvantaged. This cutpoint may not be ideal. In examining the association between neighborhood disadvantage and cortisol, we conducted a sensitivity analysis in which we changed the cutpoint from the 33rd percentile to the 25th, 20th, 15th, 10th, and 5th. Results were similar for all cutpoints except the 5th percentile, which may be due to the large loss in sample size necessary to retain positivity. While lowering the percentile of the cutpoint likely results in a

more homogenous exposure, that advantage must be balanced with disadvantages of loss of sample size and loss of generalizability. In addition, at a certain point, gains in homogeneity may stop being meaningful. For these reasons, we would like to evaluate and utilize recent machine learning methods to choose an optimal cutpoint [15] in future work.

Third, measurement error may be differential by urbanicity, which is a particular concern for Aim 1. This issue is typically called measurement variance, and we performed a sensitivity analysis to assess whether our results could be an artifact of measurement variance of neighborhood disadvantage across levels of urbanicity. Multiple-group confirmatory factor analysis that allowed the loading coefficients of the neighborhood disadvantage measurement model to differ by urbanicity was used to estimate factor scores using the regression method. [2] Defining neighborhood disadvantage based on these factor scores did not change our inferences.

In addition, our finding of urbanicity as an effect modifier in Aim 1 may be spurious if there are unmeasured effect modifiers of the neighborhood disadvantage-mental health relationship at the neighborhood level that were not addressed. [10] Addressing such effect modifiers that are measured was the focus of Aim 3. We are unaware of bias equations for unmeasured confounders at the neighborhood level. Developing such equations is an area for future work.

Finally, using factor scores to value neighborhood disadvantage and then using these scores in subsequent analyses as opposed fitting a one-step structural equation model propagates random measurement error, resulting in less efficient effect estimates. [1, 6]

5.3.1.2 Theoretical limitations

There is also a more philosophical limitation to our measure of neighborhood disadvantage: it is unlikely that every adolescent we classified as being "exposed to" living in a disadvantaged neighborhood experienced the same exposure. For example, for one adolescent, living in a disadvantaged neighborhood may mean being exposed to multiple sources of neighborhood violence witnessing and victimization, loud noise from traffic and transportation, and interacting with neighbors and schoolmates who are relatively homogenous in their socioeconomic makeup. For another adolescent, living in a disadvantaged neighborhood may mean being exposed to violence witnessing, but not victimization, minimal noise, and interacting with schoolmates who have a relatively diverse socioeconomic makeup. In our analyses, both of these exposures are classified as disadvantaged neighborhood residence. This is a violation of the causal inference assumption, SUTVA, which assumes only one version of each treatment/exposure. In future work, I would like to focus on studies with well-defined and potentially modifiable treatments/exposures, as this would aid inference as well as the study's practical utility informing policies and programs. For example, the Moving to Opportunity (MTO) Study was discussed in Section 1.2.1. In MTO, the treatment is being given a voucher to move to a nonpoor neighborhood. The voucher is the same for all participants and is a specific policy intervention (albeit an unsustainable and politically unpopular one). (It is important to note that if we consider the neighborhood as the exposure instead of voucher assignment, we may again have a SUTVA problem.) While well-defined exposures such as the MTO vouchers meet the treatment consistency requirement of SUTVA, the no-interference requirement may be violated. For example, families who did not receive a voucher could nonetheless be

affected by the program when their neighbors with vouchers moved. [32] However, an emerging area of research is in developing methods to relax this assumption. [36, 38]

In addition, propensity score theory assumes that variables are measured without error. It is likely that there was at least some error in several of the variables we included in our propensity score models. There has been little work done on the effects of estimating propensity scores when the covariates are measured with error, but preliminary studies suggest that measurement error compromises the bias-reduction potential of propensity scores, especially in the case of differential measurement error. [23, 33] There has also been little work on ways to relax this assumption. [21, 24, 35] This will be one area of study in my postdoctoral work.

5.3.2 Error in measuring cortisol

Limitations in measuring cortisol levels in Aim 2 can be thought of as falling into one of two categories: limitations in the actual measurement of salivary cortisol and limitations in using salivary cortisol as a proxy for stress.

5.3.2.1 Limitations in the measurement of salivary cortisol

One source of error in the actual measurement of salivary cortisol stems from cortisol's sensitivity to a wide range of factors that were not held constant in its collection, which we discussed in Section 5.2.2. To reduce this source of error and allow for the measurement of specific dimensions of the stress response system via cortisol, controlled laboratory conditions that hold time, day of the week, and certain environmental variables (e.g., light, temperature, noise) and behaviors (e.g., level of physical activity, smoking, drinking) constant and the use of protocols that admin-

ister either a standardized acute stressor (e.g., a Trier stress test) or rest period are required. [13] Because this was not possible in the NCS-A, we cannot infer that the cortisol outcomes in Aim 2 map onto specific HPA axis dimensions.

The second source of error is the cortisol assay. Quantification of cortisol levels was done by a radioimmunoassay (Siemens Diagnostic). The sensitivity of the assay was 0.0165 ng/mL. Intra- and inter-assay coefficients of variation were 5.4% and 26.0%, respectively. Similar coefficients of variation for this method have been reported previously, [42] but the interassay coefficient of variation is large.

5.3.2.2 Limitations in using salivary cortisol as a proxy for stress

As discussed in Section 1.4.1, salivary cortisol is only one of many biomarkers that could be used to measure the multi-faceted and complex stress response system. [22] Seeman et al has argued against the use of any one measure as a proxy for allostatic load, saying "Prior research...has largely examined the role of individual biological parameters... [In contrast], allostatic load is reflected in the cumulative total of physiological dysregulations across multiple physiologic regulatory systems." [29] For example, in the MacArthur Successful Aging Study, Seeman et al used over a dozen different biologic parameters. These parameters included primary mediators, primary effects, and secondary effects and represented measures of the HPA axis (e.g., cortisol), sympathetic nervous system (e.g., norepinephrine), and effects on the metabolic (e.g., glycosylated hemoglobin), immunologic (e.g., fibrinogen) and cardiovascular systems (e.g., systolic and diastolic blood pressure). [29] Moreover, Chida and Hamer note in a 2008 review that different stress mediators show different

response patterns within stressor type. [3] So, including a variety of biomarkers can give a more accurate and complete picture of the relationship between allostatic load and an exposure of interest. Applying this to our Aim 2 results, it is possible that the relationship between neighborhood disadvantage and other indicators of allostatic load could have been significant even in the absence of a relationship with cortisol.

Although several leaders in the stress biomarker field have argued for the use of multiple measures of allostatic load, many studies continue to measure cortisol as the sole stress biomarker. This could be for the practical reason that using one indicator of allostatic load as opposed to many simplifies and reduces the cost of the study. If one chooses to measure a single indicator of allostatic load, cortisol may be a reasonable candidate, as discussed in Section 1.4.1. Free, unbound cortisol can be obtained from saliva samples whereas total—bound and unbound—cortisol can be obtained from serum samples. Unbound cortisol is likely the more relevant proxy, because it is thought to be the only component of cortisol to reach the "target tissue and elicit glucocorticoid effects." [17] Consistent with this idea, saliva cortisol has been found to better measure adrenal cortical function [40] and HPA axis activity, [9] although saliva and serum cortisol are highly correlated. [44] In addition, measuring cortisol through saliva samples is noninvasive and does not induce stress—a strength because of cortisol's sensitivity to stress. [17]

5.3.3 Error in measuring adolescent anxiety and depression

The outcome variable of emotional disorder is also subject to measurement error. However, the use of the CIDI (World Health Organization Composite International Diagnostic Interview Version 3.0) modified for adolescents 13 years and older in assess-

ing mental disorder is a key strength. The CIDI uses community survey methodology to administer standardized diagnostic interviews based on DSM criterion in a face-to-face format by lay interviewers. It is more reliable than unstandardized psychiatric diagnoses and shows good agreement with standardized psychiatric diagnoses. [43] It also has high content validity, as it is designed to correspond to DSM-IV and ICD-10 criteria. [16]

5.4 Strengths

The strengths of this dissertation are rooted in our recognition of its limitations—and our effects to minimize their influence—and in the strengths of the NCS-A dataset.

Conducting a cross-sectional, nonrandomized study challenges causal inference. However, our integration of multiple propensity score methods with regression outcome analysis addresses limitations of confounding and extrapolation due to nonpositivity. We accompany these methods with sensitivity analyses to an unobserved confounder to provide insight into the otherwise intractable limitation of unobserved confounding.

In addition, several of the covariates we used throughout this dissertation had missing data—the greatest amount of missing data was 30% for the current parental employment variable in Aim 2. We followed recommendations by Stuart et al and multiply imputed data for use in subsequent analyses. [34] Multiple imputation has been shown to require less strict assumptions than excluding those with missing data. We used multiple imputation by chained equations, which is a flexible approach in that it is nonparametric and does not assume a joint-normal distribution. [37] Fol-

lowing imputation, we looked at diagnostics within the imputed datasets, including checking convergence and comparing the imputed versus observed distributions across variables. [37]

Several other strengths derive from the NCS-A dataset. With over 10,000 adolescents, the NCS-A is the largest nationally representative survey of adolescent mental health in the U.S., compiling data collected from adolescents, parents, and GIS-coded residence for an unusually large amount of information on context. Because of these attributes, we were well-positioned to examine the role of urbanicity as a potential effect modifier of the association between neighborhood disadvantage and depression/anxiety in Aim 1, and we were well-positioned to address the gaps of small sample size and racial/ethnic and geographic homogeneity in previous research on the association between neighborhood and cortisol in Aim 2.

Because data on neighborhood residence and DSM-IV disorders were available for nearly all adolescents in the NCS-A, in Aim 1 we were able to preserve the NSC-A's sampling strengths and nationally representative interpretation by incorporating the survey design and weights into our analysis. The design-based analysis accounts for sample selection (incorporating strata and cluster variables) and nonresponse, thus addressing clustering (which was low—3 adolescents per neighborhood, on average) of adolescents within neighborhoods.

However, data on cortisol levels were only available for a subset (approximately one-fourth) of the NCS-A sample. No weights had been calculated for this subset, so we were unsure whether or not our results from Aim 2 would generalize to U.S. adolescents. Examining the potential transportability of Aim 2's findings was the motivation for Aim 3. First, we examined one-dimensional summary measures comparing the NCS-A subset with cortisol measures to those without and found few

differences. We then applied the TMLE methods evaluated in Aim 3's simulation study to estimate the marginal association between neighborhood disadvantage and cortisol slope adjusting for bias due to nonrandom treatment assignment and nonrandom sample selection. Our inferences were unchanged. In addition to applying these methods to Aim 2's research question, we provide an evaluation of multiple methods and a tutorial (including code) for applying the recommended TMLE implementations. In doing so, our aim is to communicate to applied epidemiologic researchers the importance of addressing transportability and the ease of doing so using TMLE.

5.5 Conclusion

In conclusion, this dissertation contributes to the evidence base that neighborhood influences stress and mental health. We found that living in a disadvantaged neighborhood is associated with current depression and anxiety in adolescents if that neighborhood is in an urban area, but not otherwise. The neighborhood environment may affect mental health through dysregulation of the stress response system. We found evidence that living in a disadvantaged neighborhood influences adolescent cortisol levels during a novel interview situation. This association was maintained when we adjusted for nonrandom selection into the cortisol sample. Because the challenge of adjusting for both treatment and selection bias in the presence of effect heterogeneity has not been extensively researched, we conducted a simulation study to inform our choice of analysis method and provide a tutorial for applied researchers on how to implement the method.

Appendix A

Future Related Work

A.1 Relationship between cortisol and mental health

In Chapter 2, we examined the association between neighborhood disadvantage and anxiety/depression, which spans our conceptual model. In Chapter 3, we examined the association between the first half of our conceptual model—neighborhood disadvantage and cortisol. It is natural, then, that in future work we examine the association between the second half of our conceptual model—cortisol and anxiety/depression.

As discussed in the Introduction, research has linked cortisol levels and other biomarkers of the stress response system to the cross-sectional association and longitudinal risk of anxiety and depression. For example, cortisol dysregulation and elevated levels are associated with depression. [7, 12, 13] Cortisol levels may remain elevated for the duration of the major depressive episode, and have been shown to reduce hippocampal volume. [10,11] This reduced hippocampal volume has also been associated with anxiety disorders like PTSD, which suggests that dysregulation of the stress response system is at least partially responsible. [1, 8]

Based on this previous research as well as our research in Aim 2, we hypothesize that higher pre-interview cortisol levels and steeper rates of cortisol decline would be associated with past-year depression and anxiety. Additionally, we hypothesize that this association is moderated by persistent and severe childhood trauma such that adolescents with exposure to such trauma will have lower pre-interview levels and flatter rates of decline, but will be more likely to have experienced past-year depression and anxiety. Therefore, we would expect the presence of trauma to attenuate the association between coritsol levels, slope and depression/anxiety.

We performed a preliminary analysis testing this hypothesis, the results of which are shown in Tables A.1 and A.2 and Figures A.1 and A.2 below. Instead of the trauma operationalization used in Aim 2, we operationalized trauma as being physically abused by either or both parents, as this was a more influential predictor. In the model with past-year depressive disorder as the outcome, we found that our hypothesis was confirmed, but did not reach the level of statistical significance. This may be in part due to the relatively low number of adolescents who report physical abuse by their parents (180 out of 1,642) and the relatively low number who were classified as having past-year depressive disorder (77 out of 1,642). Results for past-year emotional disorder were similar. In contrast, cortisol levels had very little influence on past-year anxiety disorder.

Table A.1: Pre-interview coritisol regression results. Estimates and 90% CI[1]

Variable	MDD	Anxiety	Either
log(cortisol)	0.223	0.003	0.220*
	(−0.041, 0.487)	(−0.110, 0.116)	(0.008, 0.432)
physical abuse	0.497	0.047	0.548**
	(−0.023, 1.018)	(−0.241, 0.334)	(0.146, 0.949)
sample time	−5.461	−3.921	−0.223
	(−13.753, 2.832)	(−8.704, 0.862)	(−6.896, 6.450)
sample time^2	−7.173*	1.823	−10.097**
	(−13.817, −0.530)	(−1.986, 5.632)	(−16.523, −3.670)
female	0.713**	0.636***	0.670**
	(0.260, 1.166)	(0.372, 0.900)	(0.244, 1.097)
age	0.176*	0.050	0.175**
	(0.025, 0.326)	(−0.018, 0.118)	(0.059, 0.292)
weekend	−0.449*	−0.001	−0.400*
	(−0.858, −0.040)	(−0.225, 0.223)	(−0.758, −0.042)
summer	−0.615	−0.161	−0.228
	(−1.257, 0.027)	(−0.554, 0.232)	(−0.854, 0.398)
fall	−0.419	−0.174	−0.087
	(−1.207, 0.369)	(−0.554, 0.207)	(−0.780, 0.605)
winter	−0.470	−0.461	−0.207

Continued on next page

Table A.1 – continued from previous page

Variable	MDD	Anxiety	Either
	(−1.285, 0.346)	(−0.981, 0.059)	(−0.905, 0.492)
black	−0.241	0.491**	−0.050
	(−0.830, 0.349)	(0.142, 0.840)	(−0.603, 0.503)
other race	−0.862	0.192	−0.470
	(−1.720, −0.005)	(−0.339, 0.723)	(−1.212, 0.271)
white	−0.622**	−0.284	−0.418
	(−1.117, −0.127)	(−0.595, 0.027)	(−0.991, 0.155)
log(cortisol)×abuse	−0.693*	0.034	−0.652
	(−1.299, −0.087)	(−0.294, 0.361)	(−1.339, 0.035)

Table A.2: Cortisol slope regression results. Estimates and 90% CI[2]

Variable	MDD	Anxiety	Either
cortisol slope	−1.570	0.207	−1.106
	(−3.484, 0.344)	(−0.688, 1.102)	(−3.106, 0.895)
physical abuse	0.474	0.030	0.529**
	(−0.068, 1.016)	(−0.265, 0.326)	(0.119, 0.940)
sample time	−5.558	−3.876	−0.778
	(−13.492, 2.375)	(−8.568, 0.816)	(−8.029, 6.472)
Continued on next page			

[1] *p<0.1; **p<0.05; ***p<0.01

Table A.2 – continued from previous page

Variable	MDD	Anxiety	Either
sample time^2	-7.133^{*}	1.869	-9.918^{**}
	$(-13.655, -0.611)$	$(-1.852, 5.591)$	$(-16.207, -3.630)$
female	0.689^{**}	0.636^{***}	0.651^{**}
	(0.245, 1.132)	(0.367, 0.906)	(0.229, 1.073)
age	0.178^{*}	0.048	0.178^{**}
	(0.026, 0.329)	$(-0.019, 0.115)$	(0.060, 0.296)
weekend	-0.456^{*}	0.008	-0.409^{*}
	$(-0.873, -0.038)$	$(-0.216, 0.232)$	$(-0.774, -0.045)$
summer	-0.615	-0.161	-0.234
	$(-1.242, 0.012)$	$(-0.556, 0.235)$	$(-0.854, 0.386)$
fall	-0.426	-0.167	-0.085
	$(-1.193, 0.342)$	$(-0.554, 0.221)$	$(-0.756, 0.585)$
winter	-0.461	-0.450	-0.199
	$(-1.253, 0.332)$	$(-0.978, 0.077)$	$(-0.878, 0.481)$
black	-0.232	0.497^{**}	-0.043
	$(-0.831, 0.368)$	(0.150, 0.844)	$(-0.605, 0.519)$
other race	-0.867	0.192	-0.487
	$(-1.727, -0.007)$	$(-0.331, 0.715)$	$(-1.242, 0.269)$
white	-0.617^{**}	-0.284	-0.407
	$(-1.112, -0.123)$	$(-0.595, 0.027)$	$(-0.977, 0.163)$

Continued on next page

[2] *p<0.1; **p<0.05; ***p<0.01

Table A.2 – continued from previous page

Variable	MDD	Anxiety	Either
cortisol×abuse	5.649*	−2.694	4.732
	(0.672, 10.626)	(−5.560, 0.172)	(−0.843, 10.306)

Figure A.1: Associations between pre-interview cortisol level and log odds of disorder by presence of abuse. Estimates and 95% CIs.

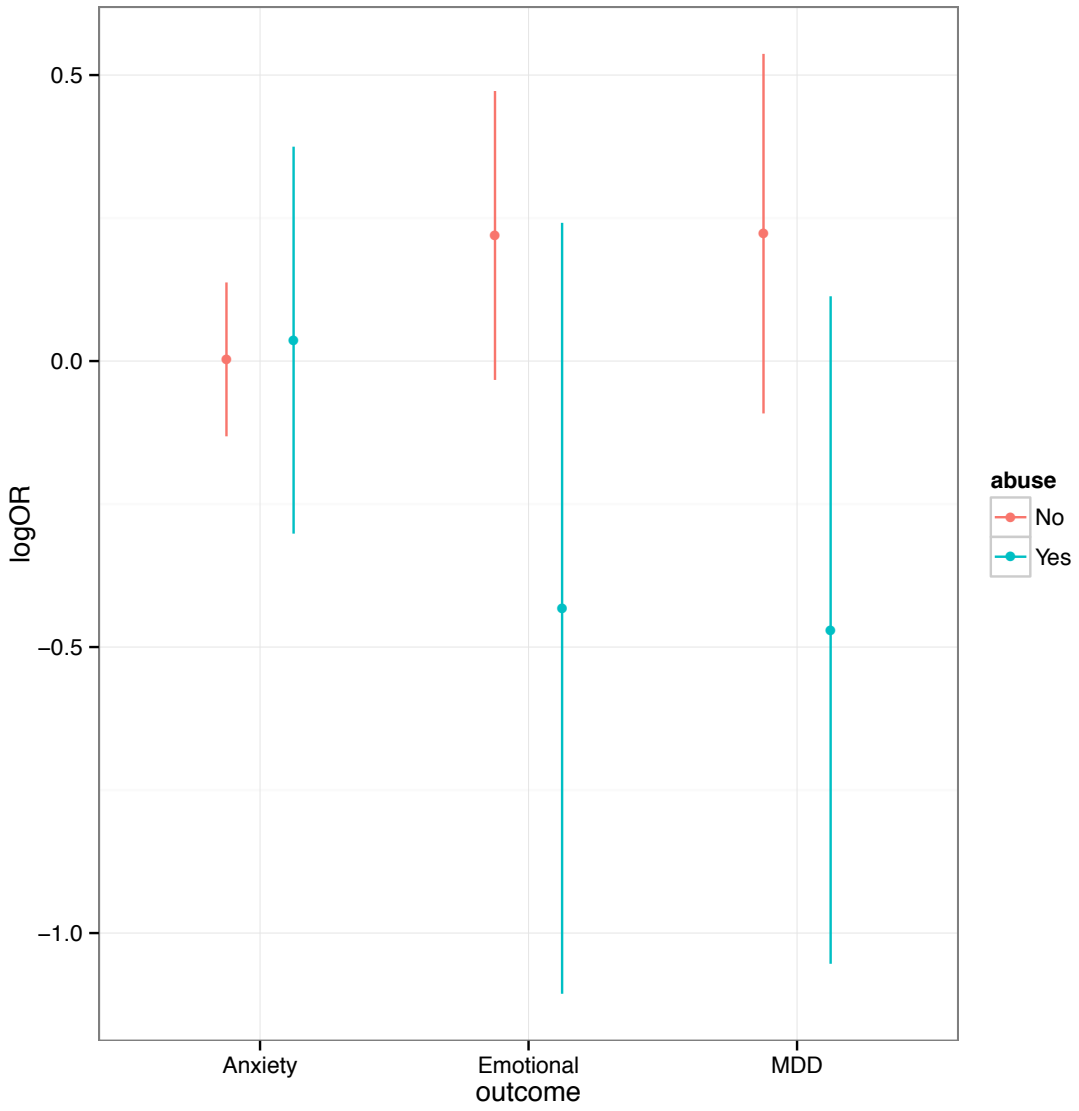
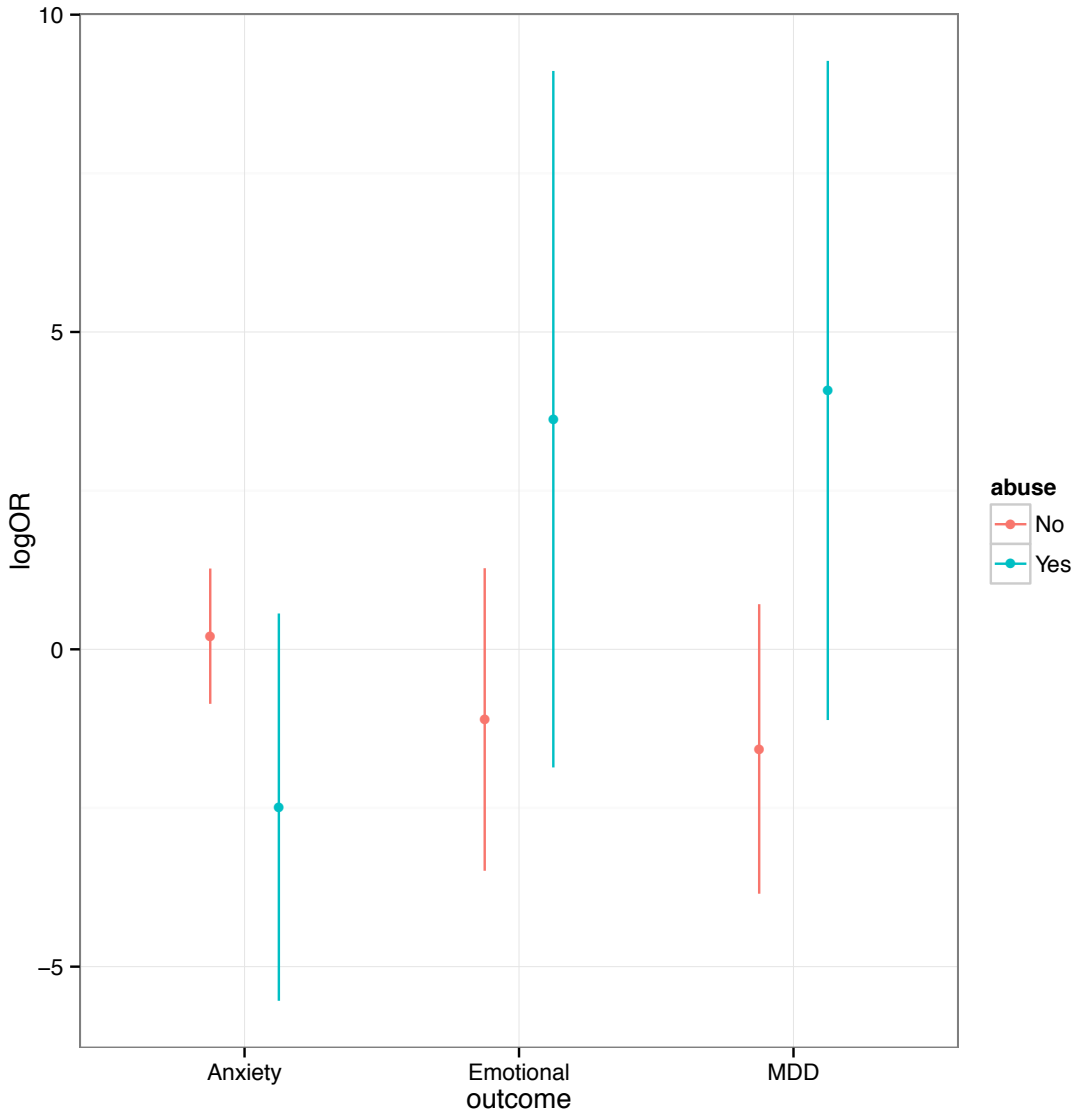


Figure A.2: Associations between cortisol slope and log odds of disorder by presence of abuse. Estimates and 95% CIs.



We performed several initial sensitivity analyses. First, we repeated the analysis restricting the dataset to samples that had been taken after 3PM and after 5PM. Results were similar but remained non-significant, and sample size diminished markedly for the after 5PM subset. We also repeated the analysis using predicted cortisol values at a given sampling time. Results changed appreciably towards the null using these

predicted values.

We caution that these results are preliminary. For a more thorough analysis, we would need to apply methods that would balance potential powerful confounders such as cortisol sampling time and weekend versus weekday, as we did in Aim 2. In this case, using propensity score methods to do so would be possible, but more complex, as our exposure of interest is continuous. [5] Relatedly, we would expect more practical positivity violations. We are also limited by the noise in the cortisol samples, which have an interassay coefficient of variation of 26.0%. Examining a relationship between cortisol and mental health may be a scenario where the design of a highly controlled experiment is important to reduce noise in order to detect an association with rare outcomes like prevalent major depressive disorder or anxiety. In addition, the design of such an experiment should include more detailed measures on childhood traumas in order to estimate these subgroup effects.

A.2 Buffering of the neighborhood disadvantage-cortisol relationship by religion

In Chapter 3, we examined the association between neighborhood disadvantage and cortisol. We hypothesized that living in a disadvantaged neighborhood would entail exposure to stressors, which would dysregulate adolescent's stress response system. It is plausible that factors such as social support and coping mechanisms could act to buffer this association. [2–4,6] Religious participation may influence both social support and coping, and there is limited evidence that spirituality (which we are aware is different from religious affiliation) can act to calm HPA axis components

such as cortisol. [9]

We tested the possible buffering influence of religion on cortisol levels by categorizing adolescents based on whether they self-associated with a relgion or not and including an interaction term between religion and neighborhood disadvantage status. The results are shown in Tables A.3 and A.4 and Figure A.3. As hypothesized, we see that affiliation with a religion dampens the association between living in a disadvantaged neighborhood and higher pre-interview cortisol levels and also dampens the association between living in a disadvantaged neighborhood and steeper cortisol slope. The interaction between neighborhood disadvantage and religion is statistically significant at the 0.05 level when cortisol slope is the outcome, but does not reach statistical significance when pre-interview cortisol levels is the outcome.

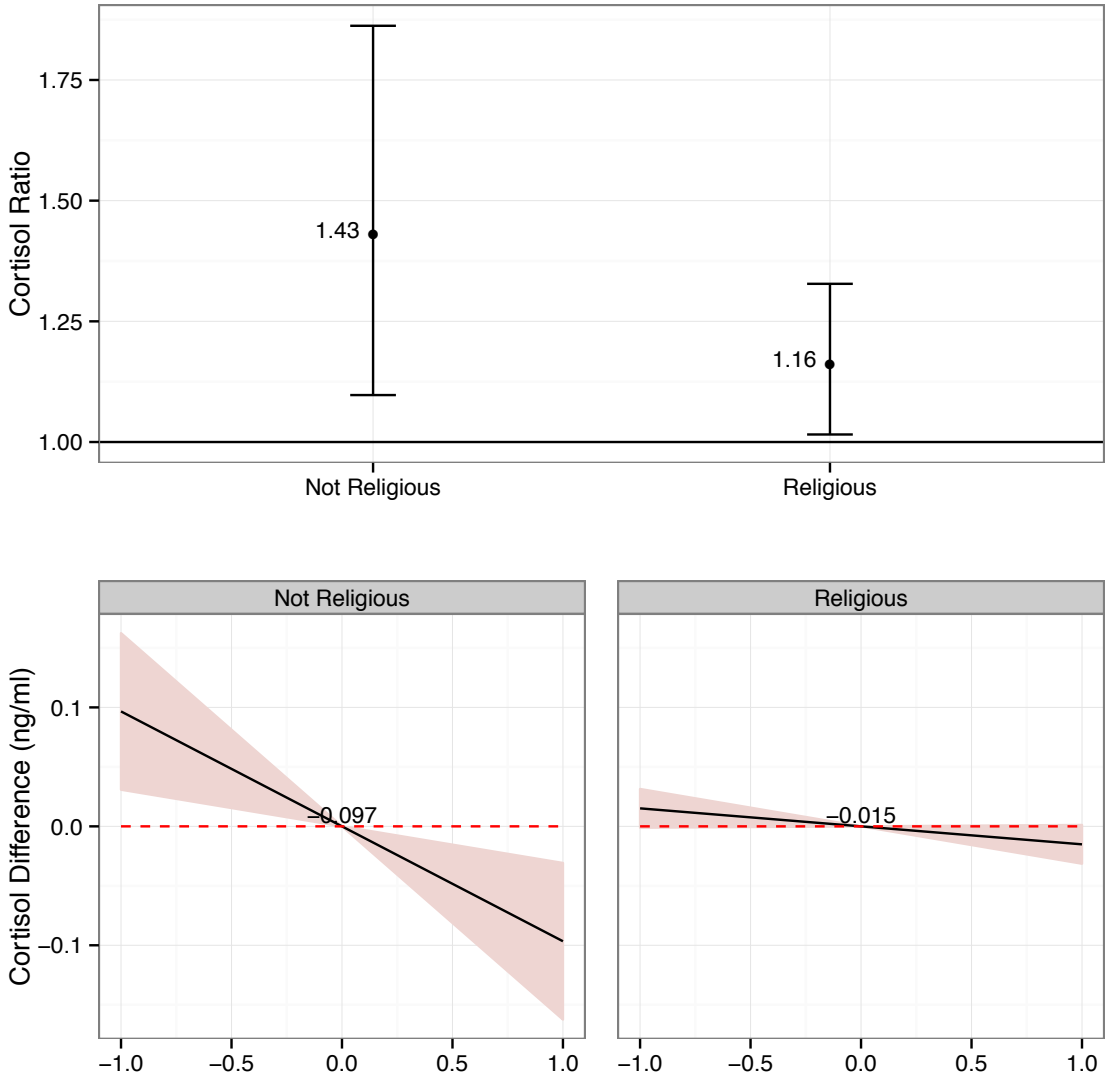
Table A.3: Pre-interview cortisol regression results. Estimates and 95% CIs.

	Coefficient	(lower	upper)
Neighborhood disadvantage	0.357	0.093	0.622
Religious affiliation	0.081	-0.172	0.334
Sampling time	-0.000	-0.000	-0.000
Female	0.008	-0.076	0.092
Age	0.039	-0.003	0.081
Urban	0.082	-0.097	0.262
Suburban	0.005	-0.158	0.168
Maternal age	-0.003	-0.013	0.007
Maternal eduation	-0.036	-0.125	0.052
2nd gen immigrant	-0.172	-0.500	0.157
3rd or later gen immigrant	-0.128	-0.405	0.149
Midwest	-0.021	-0.255	0.212
South	0.124	-0.066	0.315
West	0.079	-0.107	0.265
Summer	0.135	-0.031	0.300
Fall	0.175	0.003	0.347
Winter	0.082	-0.237	0.401
Propensity score	-0.437	-0.772	-0.102
Neighborhood disadvantage $\times$ religion	-0.208	-0.524	0.108

Table A.4: Cortisol slope regression results. Estimates and 95% CIs.

	Coefficient	(lower	upper)
Neighborhood disadvantage	-0.097	-0.162	-0.031
Religious affiliatio	-0.056	-0.114	0.002
Sampling time	0.000	0.000	0.000
Female	-0.006	-0.032	0.020
Age	-0.002	-0.009	0.005
Urban	-0.020	-0.041	0.002
Suburban	-0.010	-0.025	0.005
Maternal age	0.001	-0.001	0.002
Maternal eduation	0.003	-0.010	0.015
Language	0.002	-0.021	0.025
2nd gen immigrant	-0.017	-0.081	0.046
3rd or later gen immigrant	-0.017	-0.080	0.047
Citizen	0.045	-0.006	0.095
Midwest	0.016	-0.036	0.068
South	-0.002	-0.045	0.041
West	0.017	-0.029	0.062
Summer	0.009	-0.018	0.036
Fall	0.007	-0.019	0.034
Winter	0.014	-0.020	0.047
Propensity score	0.042	-0.006	0.090
Neighborhood disadvantage $\times$ religion	0.081	0.011	0.152
Female $\times$ age	0.000	-0.011	0.011

Figure A.3: Conditional expected ratios of pre-interview cortisol levels and conditional expected differences in cortisol slope during the late decline period comparing adolescents living in disadvantaged versus non-disadvantaged neighborhoods who are religious and not religious using Adjusted Model 1 from Aim 2. Top row: Ratios of point-in-time pre-interview cortisol levels. Error bars represent 95 CI for the mean. Bottom row: Differences in cortisol slope. Shaded areas represent 95 CI for the mean.



In future work, we plan to supplement this preliminary analysis by exploring the potential buffering effects of other factors such as social support, involvement in school/community activities or sports, and coping strategies. Together, these anal-

yses will identify modifiable factors that may promote resiliency for youth living in disadvantaged neighborhoods.

References for Chapter 1

[1] Owhofasa O Agbedia, Vijay R Varma, Christopher L Seplaki, Teresa E Seeman, Linda P Fried, Lingsheng Li, Gregory C Harris, George W Rebok, Qian-Li Xue, Erwin J Tan, et al. Blunted diurnal decline of cortisol among older adults with low socioeconomic status. *Annals of the New York Academy of Sciences*, 1231(1):56–64, 2011.

[2] Jennifer Ahern and Sandro Galea. Collective efficacy and major depression in urban neighborhoods. *American journal of epidemiology*, 173(12):1453–1462, 2011.

[3] Carol S Aneshensel, Richard G Wight, Dana Miller-Martinez, Amanda L Botticello, Arun S Karlamangla, and Teresa E Seeman. Urban neighborhoods and depressive symptoms among older adults. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 62(1):S52–S59, 2007.

[4] John R Beard, Magda Cerdá, Shannon Blaney, Jennifer Ahern, David Vlahov, and Sandro Galea. Neighborhood characteristics and change in depressive symptoms among older residents of new york city. *Journal Information*, 99(7), 2009.

[5] Lisa F Berkman. Introduction: seeing the forest and the treesâĂŤfrom obser-

vation to experiments in social epidemiology. *Epidemiologic Reviews*, 26(1):2–6, 2004.

[6] Katherine Bevans, Arleen Cerbone, and Stacy Overstreet. Relations between recurrent trauma exposure and recent life stress and salivary cortisol among children. *Development and psychopathology*, 20(1):257, 2008.

[7] Chloe E Bird, Teresa Seeman, José J Escarce, Ricardo Basurto-Dávila, Brian K Finch, Tamara Dubowitz, Melonie Heron, Lauren Hale, Sharon Stein Merkin, Margaret Weden, et al. Neighbourhood socioeconomic status and biological âĂŸwear and tearâĂŹin a nationally representative sample of us adults. *Journal of Epidemiology and Community Health*, 64(10):860–865, 2010.

[8] Emma Robertson Blackmore, Stephen A Stansfeld, Iris Weller, Sarah Munce, Brandon M Zagorski, and Donna E Stewart. Major depressive episodes and work stress: results from a national population survey. *American Journal of Public Health*, 97(11):2088–2093, 2007.

[9] Amanda L Botticello. A multilevel analysis of gender differences in psychological distress over time. *Journal of Research on Adolescence*, 19(2):217–247, 2009.

[10] Allison B Brenner, Marc A Zimmerman, Jose A Bauermeister, and Cleopatra H Caldwell. The physiological expression of living in disadvantaged neighborhoods for youth. *Journal of youth and adolescence*, pages 1–15, 2013.

[11] Joshua Breslau, Michael Lane, Nancy Sampson, and Ronald C Kessler. Mental disorders and subsequent educational attainment in a us national sample. *Journal of psychiatric research*, 42(9):708–716, 2008.

[12] George W Brown, Bernice Andrews, Antonia T Bifulco, and HO Veiel. Self esteem and depression: I. measurement issues and prediction of onset. *Social Psychiatry and Psychiatric Epidemiology*, 1990.

[13] Stephanie Campbell, Michael Marriott, Claude Nahmias, and Glenda M MacQueen. Lower hippocampal volume in patients suffering from depression: a meta-analysis. *American Journal of Psychiatry*, 161(4):598–607, 2004.

[14] John Cassel. The contribution of the social environment to host resistance. *American journal of epidemiology*, 104(2):107–123, 1976.

[15] Margaret OâĂŹBrien Caughy, Saundra Murray Nettles, and Patricia J OâĂŹCampo. The effect of residential neighborhood on child behavior problems in first grade. *American Journal of Community Psychology*, 42(1-2):39–50, 2008.

[16] Edith Chen and Laurel Q Paterson. Neighborhood, family, and subjective socioeconomic status: How do they relate to adolescent health? *Health Psychology*, 25(6):704, 2006.

[17] Cheryl Clark, Louise Ryan, Ichiro Kawachi, Marina J Canner, Lisa Berkman, and Rosalind J Wright. Witnessing community violence in residential neighborhoods: a mental health hazard for urban women. *Journal of Urban Health*, 85(1):22–38, 2008.

[18] Sheldon Cohen. Psychosocial models of the role of social support in the etiology of physical disease. *Health psychology*, 7(3):269, 1988.

[19] Sheldon Cohen, William J Doyle, and Andrew Baum. Socioeconomic status is associated with stress hormones. *Psychosomatic Medicine*, 68(3):414–420, 2006.

[20] Sheldon Cohen, Joseph E Schwartz, Elissa Epel, Clemens Kirschbaum, Steve Sidney, and Teresa Seeman. Socioeconomic status, race, and diurnal cortisol decline in the coronary artery risk development in young adults (cardia) study. *Psychosomatic Medicine*, 68(1):41–50, 2006.

[21] Carolyn E Cutrona, Daniel W Russell, P Adama Brown, Lee Anna Clark, Robert M Hessling, and Kelli A Gardner. Neighborhood context, personality, and stressful life events as predictors of depression among african american women. *Journal of abnormal psychology*, 114(1):3, 2005.

[22] Carolyn E Cutrona, Daniel W Russell, Robert M Hessling, P Adama Brown, and Velma Murry. Direct and moderating effects of community context on the psychological well-being of african american women. *Journal of Personality and Social Psychology; Journal of Personality and Social Psychology*, 79(6):1088, 2000.

[23] Sally S Dickerson and Margaret E Kemeny. Acute stressors and cortisol responses: a theoretical integration and synthesis of laboratory research. *Psychological bulletin*, 130(3):355, 2004.

[24] Ana V Diez Roux. Estimating neighborhood health effects: the challenges of causal inference in a complex world. *Social science & medicine*, 58(10):1953–1960, 2004.

[25] Ana V Diez Roux and Christina Mair. Neighborhoods and health. *Annals of the New York Academy of Sciences*, 1186(1):125–145, 2010.

[26] D Phuong Do, Ana V Diez Roux, Anjum Hajat, Amy H Auchincloss, Sharon Stein Merkin, Nalini Ranjit, Steven Shea, and Teresa Seeman. Circadian

rhythm of cortisol and neighborhood characteristics in a population-based sample: The multi-ethnic study of atherosclerosis. *Health & place*, 17(2):625–632, 2011.

[27] Rebecca C Fauth, Tama Leventhal, and Jeanne Brooks-Gunn. Short-term effects of moving from public housing in poor to middle-class neighborhoods on low-income, minority adultsâĂŹ outcomes. *Social science & medicine*, 59(11):2271–2284, 2004.

[28] Rebecca C Fauth, Tama Leventhal, and Jeanne Brooks-Gunn. Seven years later: Effects of a neighborhood mobility program on poor black and latino adults' well-being. *Journal of health and social behavior*, 49(2):119–130, 2008.

[29] Lia CH Fernald and Megan R Gunnar. Poverty-alleviation program participation and salivary cortisol in very low-income children. *Social science & medicine*, 68(12):2180–2189, 2009.

[30] Robert Finlay-Jones and George W Brown. Types of stressful life event and the onset of anxiety and depressive disorders. *Psychological Medicine*, 11(4):803–815, 1981.

[31] Kevin M Fitzpatrick, Darlene R Wright, Bettina F Piko, and Mark LaGory. Depressive symptomatology, exposure to violence, and the role of social capital among african american adolescents. *American Journal of Orthopsychiatry*, 75(2):262–274, 2005.

[32] Holly Foster and Jeanne Brooks-Gunn. Toward a stress process model of childrenâĂŹs exposure to physical family and community violence. *Clinical child and family psychology review*, 12(2):71–94, 2009.

[33] Sandro Galea, Jennifer Ahern, Arijit Nandi, Melissa Tracy, John Beard, and David Vlahov. Urban neighborhood poverty and the incidence of depression in a population-based cohort study. *Annals of epidemiology*, 17(3):171–179, 2007.

[34] Sandro Galea, Jennifer Ahern, Sasha Rudenstine, Zachary Wallace, and David Vlahov. Urban built environment and depression: a multilevel analysis. *Journal of Epidemiology and Community Health*, 59(10):822–827, 2005.

[35] Tiffany L Gary-Webb, Kesha Baptiste-Roberts, Luu Pham, Jacqueline Wesche-Thobaben, Jennifer Patricio, F Xavier Pi-Sunyer, Arleen F Brown, LaShanda Jones-Corneille, Frederick L Brancati, et al. Neighborhood socioeconomic status, depression, and health status in the look ahead (action for health in diabetes) study. *BMC public health*, 11(1):349, 2011.

[36] Arline T Geronimus. The weathering hypothesis and the health of african-american women and infants: evidence and speculations. *Ethnicity & disease*, 2(3):207–221, 1991.

[37] Sherry Glied and Daniel S Pine. Consequences and correlates of adolescent depression. *Archives of pediatrics & adolescent medicine*, 156(10):1009, 2002.

[38] Isabelle Godin, France Kittel, Yves Coppieters, and Johannes Siegrist. A prospective study of cumulative job stress in relation to mental health. *BMC Public Health*, 5(1):67, 2005.

[39] Nancy A Gonzales, Stefany Coxe, Mark W Roosa, Rebecca MB White, George P Knight, Katharine H Zeiders, and Delia Saenz. Economic hardship, neighborhood context, and parenting: Prospective effects on mexican–american

adolescentâĂŹs mental health. *American journal of community psychology*, 47(1-2):98–113, 2011.

[40] Per E Gustafsson, Henrik Anckarsäter, Paul Lichtenstein, Nina Nelson, and Per A Gustafsson. Does quantity have a quality all its own? cumulative adversity and up-and down-regulation of circadian salivary cortisol levels in healthy children. *Psychoneuroendocrinology*, 35(9):1410–1415, 2010.

[41] Leslie Morrison Gutman and Arnold J Sameroff. Continuities in depression from adolescence to young adulthood: Contrasting ecological influences. *Development and psychopathology*, 16(04):967–984, 2004.

[42] Daniel A Hackman, Laura M Betancourt, Nancy L Brodsky, Hallam Hurt, and Martha J Farah. Neighborhood disadvantage and adolescent stress reactivity. *Frontiers in human neuroscience*, 6, 2012.

[43] Anjum Hajat, Ana Diez-Roux, Tracy G Franklin, Teresa Seeman, Sandi Shrager, Nalini Ranjit, Cecilia Castro, Karol Watson, Brisa Sanchez, and Clemens Kirschbaum. Socioeconomic and race/ethnic differences in daily salivary cortisol profiles: The multi-ethnic study of atherosclerosis. *Psychoneuroendocrinology*, 35(6):932–943, 2010.

[44] Phillip L Hammack, Maryse H Richards, Zupei Luo, Emily S Edlynn, and Kevin Roy. Social support factors as moderators of community violence exposure among inner-city african american young adolescents. *Journal of Clinical Child and Adolescent Psychology*, 33(3):450–462, 2004.

[45] David J Harding. Counterfactual models of neighborhood effects: The effect

of neighborhood poverty on dropping out and teenage pregnancy1. *American Journal of Sociology*, 109(3):676–719, 2003.

[46] Claire Henderson, Ana V Diez Roux, David R Jacobs, Catarina I Kiefe, Delia West, and David R Williams. Neighbourhood characteristics, individual level socioeconomic factors, and depressive symptoms in young adults: the cardia study. *Journal of Epidemiology and Community Health*, 59(4):322–328, 2005.

[47] Miguel Hernan and James Robins. *Causal inference*. Chapman Hall, CRC, 2013.

[48] Miguel A Hernán and James M Robins. Estimating causal effects from epidemiological data. *Journal of epidemiology and community health*, 60(7):578–586, 2006.

[49] Daniel E Ho, Kosuke Imai, Gary King, and Elizabeth A Stuart. Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political analysis*, 15(3):199–236, 2007.

[50] Paul W Holland. Statistics and causal inference. *Journal of the American Statistical Association*, 81(396):945–960, 1986.

[51] Daniel G Horvitz and Donovan J Thompson. A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47(260):663–685, 1952.

[52] Celia F Hybels, Dan G Blazer, Carl F Pieper, Bruce M Burchett, Judith C Hays, Gerda G Fillenbaum, Laura D Kubzansky, and Lisa F Berkman. Sociodemographic characteristics of the neighborhood and depressive symptoms

in older adults: using multilevel modeling in geriatric psychiatry. *American Journal of Geriatric Psych*, 14(6):498–506, 2006.

[53] Joanne DeSanto Iennaco, Mark R Cullen, Linda Cantley, Martin D Slade, Martha Fiellin, and Stanislav V Kasl. Effects of externally rated job demand and control on depression diagnosis claims in an industrial cohort. *American journal of epidemiology*, 171(3):303–311, 2010.

[54] Kosuke Imai, Gary King, and Elizabeth A Stuart. Misunderstandings between experimentalists and observationalists about causal inference. *Journal of the royal statistical society: series A (statistics in society)*, 171(2):481–502, 2008.

[55] Joseph DY Kang and Joseph L Schafer. Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data. *Statistical Science*, 22(4):523–539, 2007.

[56] Robert Karasek, Dean Baker, Frank Marxer, Anders Ahlbom, and Tores Theorell. Job decision latitude, job demands, and cardiovascular disease: a prospective study of swedish men. *American journal of public health*, 71(7):694–705, 1981.

[57] Rebecca A Karb, Michael R Elliott, Jennifer B Dowd, and Jeffrey D Morenoff. Neighborhood-level stressors, social support, and diurnal patterns of cortisol: The chicago community adult health study. *Social science & medicine*, 75(6):1038–1047, 2012.

[58] Lawrence F Katz, Jeffrey R Kling, and Jeffrey B Liebman. Moving to opportunity in boston: Early results of a randomized mobility experiment. *The Quarterly Journal of Economics*, 116(2):607–654, 2001.

[59] Shona J Kelly, Robert Young, Helen Sweeting, Joachim E Fischer, and Patrick West. Levels and confounders of morning cortisol collected from adolescents in a naturalistic (school) setting. *Psychoneuroendocrinology*, 33(9):1257–1268, 2008.

[60] R Kessler, Sergio Aguilar-Gaxiola, Jordi Alonso, Somnath Chatterji, Sing Lee, Johan Ormel, T Üstün, and P Wang. Special articles. the global burden of mental disorders: an update from the who world mental health (wmh) surveys. *Epidemiologia e psichiatria sociale*, 18(1):23, 2009.

[61] Ronald C Kessler, Hagop S Akiskal, Minnie Ames, Howard Birnbaum, Paul Greenberg, Robert MA Hirschfeld, Robert Jin, Kathleen R Merikangas, and Philip S Wang. The prevalence and effects of mood disorders on work performance in a nationally representative sample of us workers. *The American journal of psychiatry*, 163(9):1561, 2006.

[62] Ronald C Kessler, Shelli Avenevoli, E Jane Costello, Jennifer Greif Green, Michael J Gruber, Steven Heeringa, Kathleen R Merikangas, Beth-Ellen Pennell, Nancy A Sampson, and Alan M Zaslavsky. Design and field procedures in the us national comorbidity survey replication adolescent supplement (ncs-a). *International Journal of Methods in Psychiatric Research*, 18(2):69–83, 2009.

[63] Ronald C Kessler, Shelli Avenevoli, E Jane Costello, Jennifer Greif Green, Michael J Gruber, Steven Heeringa, Kathleen R Merikangas, Beth-Ellen Pennell, Nancy A Sampson, and Alan M Zaslavsky. National comorbidity survey replication adolescent supplement (NCS-A): II. overview and design. *Journal of the American Academy of Child & Adolescent Psychiatry*, 48(4):380–385, 2009.

[64] Ronald C Kessler, Shelli Avenevoli, Jennifer Green, Michael J Gruber, Margaret Guyer, Yulei He, Robert Jin, Joan Kaufman, Nancy A Sampson, Alan M Zaslavsky, et al. National comorbidity survey replication adolescent supplement (NCS-A): III. concordance of DSM-IV CIDI diagnoses with clinical reassessments. *Journal of the American Academy of Child & Adolescent Psychiatry*, 48(4):386–399, 2009.

[65] Ronald C Kessler, Steven Heeringa, Matthew D Lakoma, Maria Petukhova, Agnes E Rupp, Michael Schoenbaum, Philip S Wang, and Alan M Zaslavsky. The individual-level and societal-level effects of mental disorders on earnings in the united states: Results from the national comorbidity survey replication. *The American journal of psychiatry*, 165(6):703, 2008.

[66] Ronald C Kessler, Ellen E Walters, and Melinda S Forthofer. The social consequences of psychiatric disorders, iii: probability of marital stability. *American Journal of Psychiatry*, 155(8):1092–1096, 1998.

[67] Daniel Kim. Blues from the neighborhood? neighborhood characteristics and depression. *Epidemiologic Reviews*, 30(1):101–117, 2008.

[68] Joongbaeck Kim. Neighborhood disadvantage and mental health: The role of neighborhood disorder and social relationships. *Social Science Research*, 39(2):260–271, 2010.

[69] Clemens Kirschbaum, Karl-martin Pirke, and Dirk H Hellhammer. Preliminary evidence for reduced cortisol responsivity to psychological stress in women using oral contraceptive medication. *Psychoneuroendocrinology*, 20(5):509–514, 1995.

[70] Jeffrey R Kling, Jeffrey B Liebman, and Lawrence F Katz. Experimental analysis of neighborhood effects. *Econometrica*, 75(1):83–119, 2007.

[71] Daniel J Kruger, Thomas M Reischl, and Gilbert C Gee. Neighborhood social conditions mediate the association between physical deterioration and mental health. *American journal of community psychology*, 40(3-4):261–271, 2007.

[72] Laura D Kubzansky, SV Subramanian, Ichiro Kawachi, Martha E Fay, Mah-J Soobader, and Lisa F Berkman. Neighborhood contextual influences on depressive symptoms in the elderly. *American Journal of Epidemiology*, 162(3):253–260, 2005.

[73] Hannah Kuper, Archana Singh-Manoux, Johannes Siegrist, and Michael Marmot. When reciprocity fails: effort–reward imbalance in relation to coronary heart disease and health functioning within the whitehall ii study. *Occupational and Environmental Medicine*, 59(11):777–784, 2002.

[74] Cinira Leal, Kathy Bean, Frédérique Thomas, and Basile Chaix. Are associations between neighborhood socioeconomic characteristics and body mass index or waist circumference based on model extrapolations? *Epidemiology*, 22(5):694–703, 2011.

[75] Brian K Lee, Thomas A Glass, Matthew J McAtee, Gary S Wand, Karen Bandeen-Roche, Karen I Bolla, and Brian S Schwartz. Associations of salivary cortisol with cognitive function in the baltimore memory study. *Archives of General Psychiatry*, 64(7):810, 2007.

[76] Tama Leventhal and Jeanne Brooks-Gunn. The neighborhoods they live in: the

effects of neighborhood residence on child and adolescent outcomes. *Psychological bulletin*, 126(2):309, 2000.

[77] Tama Leventhal and Jeanne Brooks-Gunn. Moving to opportunity: an experimental study of neighborhood effects on mental health. *American Journal of Public Health*, 93(9):1576–1582, 2003.

[78] Jonas Lofors, Vania Ramírez-León, and Kristina Sundquist. Neighbourhood income and anxiety: a study based on random samples of the swedish population. *The European Journal of Public Health*, 16(6):633–639, 2006.

[79] Christina Mair, Ana V Diez Roux, and Jeffrey D Morenoff. Neighborhood stressors and social support as predictors of depressive symptoms in the chicago community adult health study. *Health & place*, 16(5):811–819, 2010.

[80] Christina Mair, Ana V Diez Roux, Mingwu Shen, Steven Shea, Theresa Seeman, Sandra Echeverria, and Ellen S O'meara. Cross-sectional and longitudinal associations of neighborhood cohesion and stressors with depressive symptoms in the multiethnic study of atherosclerosis. *Annals of epidemiology*, 19(1):49–57, 2009.

[81] Christina Mair, AV Diez Roux, and Sandro Galea. Are neighbourhood characteristics associated with depressive symptoms? a review of evidence. *Journal of Epidemiology and Community Health*, 62(11):940–946, 2008.

[82] Christina Gamache Martin, Jacqueline Bruce, and Philip A Fisher. Racial and ethnic differences in diurnal cortisol rhythms in preadolescents: The role of parental psychosocial risk and monitoring. *Hormones and behavior*, 61(5):661–668, 2012.

[83] Flora I Matheson, Rahim Moineddin, James R Dunn, Maria Isabella Creatore, Piotr Gozdyra, and Richard H Glazier. Urban neighborhoods, chronic stress, gender and depression. *Social science & medicine*, 63(10):2604–2616, 2006.

[84] Bruce S McEwen. Physiology and neurobiology of stress and adaptation: central role of the brain. *Physiological reviews*, 87(3):873–904, 2007.

[85] Bruce S McEwen and Teresa Seeman. Protective and damaging effects of mediators of stress: elaborating and testing the concepts of allostasis and allostatic load. *Annals of the New York Academy of Sciences*, 896(1):30–47, 1999.

[86] Michael J Meaney, Beth Tannenbaum, Darlene Francis, Seema Bhatnagar, Nola Shanks, Victor Viau, Dajan O'Donnell, and Paul M Plotsky. Early environmental programming hypothalamic-pituitary-adrenal responses to stress. In *Seminars in Neuroscience*, volume 6, pages 247–259. Elsevier, 1994.

[87] Maria Melchior, Lisa F Berkman, Isabelle Niedhammer, Marie Zins, and Marcel Goldberg. The mental health effects of multiple work and family demands. *Social psychiatry and psychiatric epidemiology*, 42(7):573–582, 2007.

[88] Maria Melchior, Avshalom Caspi, Barry J Milne, Andrea Danese, Richie Poulton, and Terrie E Moffitt. Work stress precipitates depression and anxiety in young, working women and men. *Psychological medicine*, 37(08):1119–1129, 2007.

[89] Kathleen R Merikangas, Minnie Ames, Lihong Cui, Paul E Stang, T Bedirhan Ustun, Michael Von Korff, and Ronald C Kessler. The impact of comorbidity of mental and physical conditions on role disability in the us adult household population. *Archives of general psychiatry*, 64(10):1180, 2007.

[90] Kathleen R Merikangas, Shelli Avenevoli, E Jane Costello, Doreen Koretz, and Ronald C Kessler. National comorbidity survey replication adolescent supplement (NCS-A): I. background and measures. *Journal of the American Academy of Child & Adolescent Psychiatry*, 48(4):367–379, 2009.

[91] Kathleen Ries Merikangas, Jian-ping He, Marcy Burstein, Sonja A Swanson, Shelli Avenevoli, Lihong Cui, Corina Benjet, Katholiki Georgiades, and Joel Swendsen. Lifetime prevalence of mental disorders in us adolescents: results from the national comorbidity survey replication–adolescent supplement (ncs-a). *Journal of the American Academy of Child & Adolescent Psychiatry*, 49(10):980–989, 2010.

[92] Anne Moses. Exposure to violence, depression, and hostility in a sample of inner city high school youth. *Journal of Adolescence*, 22(1):21–32, 1999.

[93] Carles Muntaner, William W Eaton, and Chamberlain C Diala. Social inequalities in mental health: A review of concepts and underlying assumptions. *Health:*, 4(1):89–113, 2000.

[94] Aydin Nazmi, Ana Diez Roux, Nalini Ranjit, Teresa E Seeman, and Nancy S Jenny. Cross-sectional and longitudinal associations of neighborhood characteristics with inflammatory markers: findings from the multi-ethnic study of atherosclerosis. *Health & place*, 16(6):1104–1112, 2010.

[95] J Michael Oakes. The (mis) estimation of neighborhood effects: causal inference for a practicable social epidemiology. *Social science & medicine*, 58(10):1929–1952, 2004.

[96] US Dept of Justice and United States of America. Report of the surgeon general's conference on children's mental health: A national action agenda. 2001.

[97] Glenn V Ostir, Karl Eschbach, Kyriakos S Markides, and James S Goodwin. Neighbourhood composition and depressive symptoms among older mexican americans. *Journal of Epidemiology and Community Health*, 57(12):987–992, 2003.

[98] Beverley E Pearson Murphy. Treatment of major depression with steroid suppressive drugs. *The Journal of steroid biochemistry and molecular biology*, 39(2):239–244, 1991.

[99] Karen L Petersen, Anna L Marsland, Janine Flory, Elizabeth Votruba-Drzal, Matthew F Muldoon, and Stephen B Manuck. Community socioeconomic status is associated with circulating interleukin-6 and c-reactive protein. *Psychosomatic Medicine*, 70(6):646–652, 2008.

[100] Paul M Plotsky, KV Thrivikraman, Charles B Nemeroff, Christian Caldji, Shakti Sharma, and Michael J Meaney. Long-term consequences of neonatal rearing on central corticotropin-releasing factor systems in adult male rat offspring. *Neuropsychopharmacology*, 30(12):2192–2204, 2005.

[101] Ricardo A Pollitt, Jay S Kaufman, Kathryn M Rose, Ana V Diez-Roux, Donglin Zeng, and Gerardo Heiss. Early-life and adult socioeconomic status and inflammatory risk markers in adulthood. *European journal of epidemiology*, 22(1):55–66, 2007.

[102] Susan J Popkin, James E Rosenbaum, and Patricia M Meaden. Labor market experiences of low-income black women in middle-class suburbs: Evidence from

a survey of gautreaux program participants. *Journal of Policy Analysis and Management*, 12(3):556–573, 1993.

[103] Jens C Pruessner, Mark W Baldwin, Katarina Dedovic, Robert Renwick, Najmeh Khalili Mahani, Catherine Lord, Michael Meaney, and Sonia Lupien. Self-esteem, locus of control, hippocampal volume, and cortisol regulation in young and old adulthood. *Neuroimage*, 28(4):815–826, 2005.

[104] Yanique Redwood, Amy J Schulz, Barbara A Israel, Mieko Yoshihama, Caroline C Wang, and Marshall Kreuter. Social, economic, and political processes that create built environment inequities: perspectives from urban african americans in atlanta. *Family & community health*, 33(1):53–67, 2010.

[105] Sijmen A Reijneveld and Aart H Schene. Higher prevalence of mental disorders in socioeconomically deprived urban areas in the netherlands: community or personal disadvantage? *Journal of Epidemiology and Community Health*, 52(1):2–7, 1998.

[106] Rena L Repetti, Shelley E Taylor, and Teresa E Seeman. Risky families: family social environments and the mental and physical health of offspring. *Psychological bulletin*, 128(2):330, 2002.

[107] James Robins. A new approach to causal inference in mortality studies with a sustained exposure periodâĂŤapplication to control of the healthy worker survivor effect. *Mathematical Modelling*, 7(9):1393–1512, 1986.

[108] James M Robins, Andrea Rotnitzky, and Lue Ping Zhao. Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association*, 89(427):846–866, 1994.

[109] James M Robins, Andrea Rotnitzky, and Lue Ping Zhao. Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association*, 90(429):106–121, 1995.

[110] Paul R Rosenbaum and Donald B Rubin. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55, 1983.

[111] Morris Rosenberg and Leonard I Pearlin. Social class and self-esteem among children and adults. *American Journal of Sociology*, pages 53–77, 1978.

[112] Catherine E Ross and John Mirowsky. Disorder and decay the concept and measurement of perceived neighborhood disorder. *Urban Affairs Review*, 34(3):412–432, 1999.

[113] Catherine E Ross and John Mirowsky. Neighborhood disorder, subjective alienation, and distress. *Journal of Health and Social Behavior*, 50(1):49–64, 2009.

[114] Catherine E Ross, John R Reynolds, and Karlyn J Geis. The contingent meaning of neighborhood stability for residents' psychological well-being. *American Sociological Review*, pages 581–597, 2000.

[115] Donald B Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology*, 66(5):688, 1974.

[116] Donald B Rubin. Using multivariate matched sampling and regression adjustment to control bias in observational studies. *Journal of the American Statistical Association*, 74(366a):318–328, 1979.

[117] Donald B Rubin. Comment: Which ifs have causal answers. *Journal of the American Statistical Association*, 81(396):961–962, 1986.

[118] Donald B Rubin and Neal Thomas. Combining propensity score matching with additional adjustments for prognostic covariates. *Journal of the American Statistical Association*, 95(450):573–585, 2000.

[119] David R Rubinow, Robert M Post, Robert Savard, and Philip W Gold. Cortisol hypersecretion and cognitive impairment in depression. *Archives of general psychiatry*, 41(3):279, 1984.

[120] Robert J Sampson, Stephen W Raudenbush, and Felton Earls. Neighborhoods and violent crime: A multilevel study of collective efficacy. *Science*, 277(5328):918–924, 1997.

[121] Robert M Sapolsky. The influence of social hierarchy on primate health. *Science*, 308(5722):648–652, 2005.

[122] Robert M Sapolsky, Susan C Alberts, and Jeanne Altmann. Hypercortisolism associated with social subordinance or social isolation among wild baboons. *Archives of General Psychiatry*, 54(12):1137, 1997.

[123] Daniel O Scharfstein, Andrea Rotnitzky, and James M Robins. Adjusting for nonignorable drop-out using semiparametric nonresponse models. *Journal of the American Statistical Association*, 94(448):1096–1120, 1999.

[124] Mario Schootman, Elena M Andresen, Fredric D Wolinsky, Theodore K Malmstrom, J Philip Miller, and Douglas K Miller. Neighbourhood environment and the incidence of depressive symptoms among middle-aged african americans. *Journal of epidemiology and community health*, 61(6):527–532, 2007.

[125] Mary Schwab-Stone, Chuansheng Chen, Ellen Greenberger, David Silver, Judith Lichtman, and Charlene Voyce. No safe haven ii: The effects of violence exposure on urban youth. *Journal of the American Academy of Child & Adolescent Psychiatry*, 38(4):359–367, 1999.

[126] Teresa E Seeman, Eileen Crimmins, Mei-Hua Huang, Burton Singer, Alexander Bucur, Tara Gruenewald, Lisa F Berkman, and David B Reuben. Cumulative biological risk and socio-economic differences in mortality: Macarthur studies of successful aging. *Social science & medicine*, 58(10):1985–1997, 2004.

[127] Jasjeet S Sekhon. The neyman-rubin model of causal inference and estimation via matching methods, 2008.

[128] Johannes Siegrist and Michael Marmot. Health inequalities and the psychosocial environmentâĂŤtwo scientific challenges. *Social science & medicine*, 58(8):1463–1473, 2004.

[129] Eric Silver, Edward P Mulvey, and Jeffrey W Swanson. Neighborhood structural characteristics and mental disorder: Faris and dunham revisited. *Social science & medicine*, 55(8):1457–1470, 2002.

[130] Ronald L Simons, Velma Murry, Vonnie McLoyd, Kuei-Hsiu Lin, Carolyn Cutrona, Rand D Conger, et al. Discrimination, crime, ethnic identity, and parenting as correlates of depressive symptoms among african american children: A multilevel analysis. *Development and psychopathology*, 14(2):371–393, 2002.

[131] Petros Skapinakis, Glyn Lewis, Ricardo Araya, Kelvyn Jones, and Gareth Williams. Mental health inequalities in wales, uk: multi-level investigation

of the effect of area deprivation. *The British Journal of Psychiatry*, 186(5):417–422, 2005.

[132] Mai Stafford, Mary De Silva, Stephen Stansfeld, and Michael Marmot. Neighbourhood social capital and common mental disorder: testing the link in a general population sample. *Health & place*, 14(3):394–405, 2008.

[133] Stephen A Stansfeld, Rebecca Fuhrer, Martin J Shipley, and MG Marmot. Work characteristics predict psychiatric disorder: prospective results from the whitehall ii study. *Occupational and environmental medicine*, 56(5):302–307, 1999.

[134] Andrew Steptoe, Natalie Owen, Sabine R Kunz-Ebrecht, and Lena Brydon. Loneliness and neuroendocrine, cardiovascular, and inflammatory stress responses in middle-aged men and women. *Psychoneuroendocrinology*, 29(5):593–611, 2004.

[135] Susan E Stockdale, Kenneth B Wells, Lingqi Tang, Thomas R Belin, Lily Zhang, and Cathy D Sherbourne. The importance of social context: neighborhood stressors, stress-buffering mechanisms, and alcohol, drug, and mental health disorders. *Social Science & Medicine*, 65(9):1867–1881, 2007.

[136] Marilyn Taylor. Combating the social exclusion of housing estates. *Housing Studies*, 13(6):819–832, 1998.

[137] Melissa Tracy, Frederick J Zimmerman, Sandro Galea, Elizabeth McCauley, and Ann Vander Stoep. What explains the relation between family poverty and childhood depressive symptoms? *Journal of psychiatric research*, 42(14):1163–1175, 2008.

[138] Mark J van der Laan and Daniel Rubin. Targeted maximum likelihood learning. *The International Journal of Biostatistics*, 2(1), 2006.

[139] Natasja Van Vegchel, Jan De Jonge, Hans Bosma, and Wilmar Schaufeli. Reviewing the effort–reward imbalance model: drawing up the balance of 45 empirical studies. *Social science & medicine*, 60(5):1117–1131, 2005.

[140] Gerry Veenstra. Location, location, location: contextual and compositional health effects of social capital in british columbia, canada. *Social science & medicine*, 60(9):2059–2071, 2005.

[141] NWJ Wainwright and PG Surtees. Places, people, and their physical and mental functional health. *Journal of epidemiology and community health*, 58(4):333–339, 2004.

[142] Kate Walters, Elizabeth Breeze, Paul Wilkinson, Gill M Price, Chris J Bulpitt, and Astrid Fletcher. Local area deprivation and urban-rural differences in anxiety and depression among people older than 75 years in britain. *American Journal of Public Health*, 94(10):1768–1774, 2004.

[143] S Weich, Liz Twigg, G Holt, G Lewis, and K Jones. Contextual risk factors for the common mental disorders in britain: a multilevel investigation of the effects of place. *Journal of Epidemiology and Community Health*, 57(8):616–621, 2003.

[144] Scott Weich, Liz Twigg, Glyn Lewis, and Kelvyn Jones. Geographical variation in rates of common mental disorders in britain: prospective cohort study. *The British Journal of Psychiatry*, 187(1):29–34, 2005.

[145] Joanna Wieclaw, Esben Agerbo, Preben Bo Mortensen, Hermann Burr, Finn

Tuchsen, and Jens P Bonde. Psychosocial working conditions and the risk of depression and anxiety disorders in the danish workforce. *BMC Public Health*, 8(1):280, 2008.

[146] Richard G Wight, Janet R Cummings, Arun S Karlamangla, and Carol S Aneshensel. Urban neighborhood context and change in depressive symptoms in late life. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 64(2):247–251, 2009.

[147] James Q Wilson and George L Kelling. Broken windows. *Atlantic monthly*, 249(3):29–38, 1982.

[148] William J Wilson. *When work disappears: The world of the new urban poor*. Random House Digital, Inc., 1997.

[149] Yange Xue, Tama Leventhal, Jeanne Brooks-Gunn, and Felton J Earls. Neighborhood residence and mental health problems of 5-to 11-year-olds. *Archives of General Psychiatry*, 62(5):554, 2005.

[150] Irene H Yen and George A Kaplan. Poverty area residence and changes in depression and perceived health status: evidence from the alameda county study. *International Journal of Epidemiology*, 28(1):90–94, 1999.

References for Chapter 2

[1] Adrian Angold, Myrna M Weissman, Karen John, Kathleen R Merikancas, Brigitte A Prusoff, Priya Wickramaratne, G Davis Gammon, and Virgina Warner. Parent and child reports of depressive symptoms in children at low and high risk of depression. *Journal of Child Psychology and Psychiatry*, 28(6):901–915, 1987.

[2] Kenneth A Bollen and Gerhard Arminger. Observational residuals in factor analysis and structural equation models. *Sociological methodology*, 21:235–262, 1991.

[3] Joshua Breslau, Michael Lane, Nancy Sampson, and Ronald C Kessler. Mental disorders and subsequent educational attainment in a us national sample. *Journal of psychiatric research*, 42(9):708–716, 2008.

[4] Ana V Diez Roux, Catarina I Kiefe, David R Jacobs Jr, Mary Haan, Scott A Jackson, F Javier Nieto, Catherine C Paton, and Richard Schulz. Area characteristics and individual-level socioeconomic position indicators in three population-based epidemiologic studies. *Annals of Epidemiology*, 11(6):395–405, 2001.

[5] Bruce P Dohrenwend and Barbara Snell Dohrenwend. Psychiatric disorders in urban settings. *American handbook of psychiatry*, 2:426–446, 1974.

[6] Eva H DuGoff, Megan Schuler, and Elizabeth A Stuart. Generalizing observational study results: Applying propensity score methods to complex surveys. *Health services research*, 2013.

[7] Abigail A Fagan, Emily M Wright, and Gillian M Pinchevsky. Racial/ethnic differences in the relationship between neighborhood disadvantage and adolescent substance use. *Journal of Drug Issues*, 43(1):69–84, 2013.

[8] Robert E Lee Faris and Henry Warren Dunham. Mental disorders in urban areas: an ecological study of schizophrenia and other psychoses. 1939.

[9] Sandro Galea. The urban brain: new directions in research exploring the relation between cities and mood–anxiety disorders. *Depression and anxiety*, 28(10):857–862, 2011.

[10] Andrew Gelman. Struggles with survey weighting and regression modeling. *Statistical Science*, 22(2):153–164, 2007.

[11] Sherry Glied and Daniel S Pine. Consequences and correlates of adolescent depression. *Archives of pediatrics & adolescent medicine*, 156(10):1009, 2002.

[12] Nancy A Gonzales, Stefany Coxe, Mark W Roosa, Rebecca MB White, George P Knight, Katharine H Zeiders, and Delia Saenz. Economic hardship, neighborhood context, and parenting: Prospective effects on mexican–american adolescentâĂŹs mental health. *American journal of community psychology*, 47(1-2):98–113, 2011.

[13] Jackson A Goodnight, Benjamin B Lahey, Carol A Van Hulle, Joseph L Rodgers, Paul J Rathouz, Irwin D Waldman, and Brian M D'Onofrio. A quasi-

experimental analysis of the influence of neighborhood disadvantage on child and adolescent conduct problems. *Journal of abnormal psychology*, 121(1):95, 2012.

[14] Miguel A Hernán and James M Robins. Estimating causal effects from epidemiological data. *Journal of epidemiology and community health*, 60(7):578–586, 2006.

[15] Ronald C Kessler, Shelli Avenevoli, E Jane Costello, Jennifer Greif Green, Michael J Gruber, Steven Heeringa, Kathleen R Merikangas, Beth-Ellen Pennell, Nancy A Sampson, and Alan M Zaslavsky. Design and field procedures in the us national comorbidity survey replication adolescent supplement (ncs-a). *International Journal of Methods in Psychiatric Research*, 18(2):69–83, 2009.

[16] Ronald C Kessler, Shelli Avenevoli, Jennifer Green, Michael J Gruber, Margaret Guyer, Yulei He, Robert Jin, Joan Kaufman, Nancy A Sampson, Alan M Zaslavsky, et al. National comorbidity survey replication adolescent supplement (ncs-a): Iii. concordance of< i> dsm-iv</i>/cidi diagnoses with clinical reassessments. *Journal of the American Academy of Child & Adolescent Psychiatry*, 48(4):386–399, 2009.

[17] Ronald C Kessler, Steven Heeringa, Matthew D Lakoma, Maria Petukhova, Agnes E Rupp, Michael Schoenbaum, Philip S Wang, and Alan M Zaslavsky. The individual-level and societal-level effects of mental disorders on earnings in the united states: Results from the national comorbidity survey replication. *The American journal of psychiatry*, 165(6):703, 2008.

[18] Ronald C Kessler, Ellen E Walters, and Melinda S Forthofer. The social conse-

quences of psychiatric disorders, iii: probability of marital stability. *American Journal of Psychiatry*, 155(8):1092–1096, 1998.

[19] Jeffrey R Kling, Jeffrey B Liebman, and Lawrence F Katz. Experimental analysis of neighborhood effects. *Econometrica*, 75(1):83–119, 2007.

[20] Nancy Krieger, Jarvis T Chen, Pamela D Waterman, Mah-Jabeen Soobader, SV Subramanian, and Rosa Carson. Geocoding and monitoring of us socioeconomic inequalities in mortality and cancer incidence: does the choice of area-based measure and geographic level matter? the public health disparities geocoding project. *American journal of epidemiology*, 156(5):471–482, 2002.

[21] ACK Lee and R Maheswaran. The health benefits of urban green spaces: a review of the evidence. *Journal of Public Health*, 33(2):212–222, 2011.

[22] Tama Leventhal and Jeanne Brooks-Gunn. The neighborhoods they live in: the effects of neighborhood residence on child and adolescent outcomes. *Psychological bulletin*, 126(2):309, 2000.

[23] Tama Leventhal and Jeanne Brooks-Gunn. Moving to opportunity: an experimental study of neighborhood effects on mental health. *American Journal of Public Health*, 93(9):1576–1582, 2003.

[24] Kathleen Ries Merikangas, Jian-ping He, Marcy Burstein, Sonja A Swanson, Shelli Avenevoli, Lihong Cui, Corina Benjet, Katholiki Georgiades, and Joel Swendsen. Lifetime prevalence of mental disorders in us adolescents: results from the national comorbidity survey replication–adolescent supplement (ncs-a). *Journal of the American Academy of Child & Adolescent Psychiatry*, 49(10):980–989, 2010.

[25] Elizabeth Merwin, Audrey Snyder, and Elizabeth Katz. Differential access to quality rural healthcare: professional and policy challenges. *Family & Community Health*, 29(3):186–194, 2006.

[26] Kathryn M Neckerman, Gina S Lovasi, Stephen Davies, Marnie Purciel, James Quinn, Eric Feder, Nakita Raghunath, Benjamin Wasserman, and Andrew Rundle. Disparities in urban neighborhood conditions: evidence from gis measures and field observation in new york city. *Journal of Public Health Policy*, pages S264–S285, 2009.

[27] US Dept of Justice and United States of America. Report of the surgeon general's conference on children's mental health: A national action agenda. 2001.

[28] J Peen, RA Schoevers, AT Beekman, and J Dekker. The current status of urban-rural differences in psychiatric disorders. *Acta Psychiatrica Scandinavica*, 121(2):84–93, 2010.

[29] James Robins, Mariela Sued, Quanhong Lei-Gomez, and Andrea Rotnitzky. Comment: Performance of double-robust estimators when "inverse probability" weights are highly variable. *Statistical Science*, 22(4):544–559, 2007.

[30] Paul R Rosenbaum and Donald B Rubin. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55, 1983.

[31] Ana V Diez Roux. The study of group-level factors in epidemiology: rethinking variables, study designs, and analytical approaches. *Epidemiologic Reviews*, 26(1):104–111, 2004.

[32] Ana V Diez Roux, Luisa N Borrell, Mary Haan, Sharon A Jackson, and Richard

Schultz. Neighbourhood environments and mortality in an elderly cohort: results from the cardiovascular health study. *Journal of epidemiology and community health*, 58(11):917–923, 2004.

[33] Donald B Rubin. *Multiple imputation for nonresponse in surveys*, volume 307. Wiley. com, 2009.

[34] Kara E Rudolph, Elizabeth A Stuart, Thomas A Glass, and Kathleen R Merikangas. Neighborhood disadvantage in context: the influence of urbanicity on the association between neighborhood disadvantage and adolescent emotional disorders. *Social psychiatry and psychiatric epidemiology*, pages 1–9, 2013.

[35] Ronald L Simons, Velma Murry, Vonnie McLoyd, Kuei-Hsiu Lin, Carolyn Cutrona, Rand D Conger, et al. Discrimination, crime, ethnic identity, and parenting as correlates of depressive symptoms among african american children: A multilevel analysis. *Development and psychopathology*, 14(2):371–393, 2002.

[36] Natalie Slopen, Garrett M Fitzmaurice, David R Williams, and Stephen E Gilman. Common patterns of violence experiences and depression and anxiety among adolescents. *Social psychiatry and psychiatric epidemiology*, 47(10):1591–1605, 2012.

[37] Elizabeth A Stuart, Melissa Azur, Constantine Frangakis, and Philip Leaf. Multiple imputation with large data sets: a case study of the children's mental health initiative. *American journal of epidemiology*, 169(9):1133–1139, 2009.

[38] Stef Van Buuren. Multiple imputation of discrete and continuous data by fully conditional specification. *Statistical methods in medical research*, 16(3):219–242, 2007.

[39] Tyler J VanderWeele and Onyebuchi A Arah. Bias formulas for sensitivity analysis of unmeasured confounding for general outcomes, treatments, and confounders. *Epidemiology*, 22(1):42–52, 2011.

[40] Per-Olof H Wikström, Robert J Sampson, et al. Social mechanisms of community influences on crime and pathways in criminality. *The causes of conduct disorder and serious juvenile delinquency*, pages 118–148, 2003.

[41] Hans-Ulrich Wittchen. Reliability and validity studies of the who-composite international diagnostic interview (cidi): a critical review. *Journal of psychiatric research*, 28(1):57–84, 1994.

[42] Yange Xue, Tama Leventhal, Jeanne Brooks-Gunn, and Felton J Earls. Neighborhood residence and mental health problems of 5-to 11-year-olds. *Archives of General Psychiatry*, 62(5):554, 2005.

[43] Elaine L Zanutto. A comparison of propensity score and linear regression analysis of complex survey data. *Journal of Data Science*, 4(1):67–91, 2006.

References for Chapter 3

[1] Owhofasa O Agbedia, Vijay R Varma, Christopher L Seplaki, Teresa E Seeman, Linda P Fried, Lingsheng Li, Gregory C Harris, George W Rebok, Qian-Li Xue, Erwin J Tan, et al. Blunted diurnal decline of cortisol among older adults with low socioeconomic status. *Annals of the New York Academy of Sciences*, 1231(1):56–64, 2011.

[2] Greg R Alexander, Michael D Kogan, and John H Himes. 1994–1996 us singleton birth weight percentiles for gestational age by race, hispanic origin, and gender. *Maternal and child health journal*, 3(4):225–231, 1999.

[3] Hymie Anisman and Robert M Zacharko. Depression as a consequence of inadequate neurochemical adaptation in response to stressors. *The British Journal of Psychiatry*, 1992.

[4] Katherine Bevans, Arleen Cerbone, and Stacy Overstreet. Relations between recurrent trauma exposure and recent life stress and salivary cortisol among children. *Development and psychopathology*, 20(1):257, 2008.

[5] Chloe E Bird, Teresa Seeman, José J Escarce, Ricardo Basurto-Dávila, Brian K Finch, Tamara Dubowitz, Melonie Heron, Lauren Hale, Sharon Stein Merkin, Margaret Weden, et al. Neighbourhood socioeconomic status and biological

âĂŸwear and tearâĂŹin a nationally representative sample of us adults. *Journal of Epidemiology and Community Health*, 64(10):860–865, 2010.

[6] Luisa N Borrell, James D Beck, and Gerardo Heiss. Socioeconomic disadvantage and periodontal disease: the dental atherosclerosis risk in communities study. *Journal Information*, 96(2), 2006.

[7] Nienke M Bosch, Harriëtte Riese, Sijmen A Reijneveld, Martin P Bakker, Frank C Verhulst, Johan Ormel, and Albertine J Oldehinkel. Timing matters: Long term effects of adversities from prenatal period up to adolescence on adolescentsâĂŹ cortisol stress response. the trails study. *Psychoneuroendocrinology*, 37(9):1439–1447, 2012.

[8] Esther Bouma, Harriëtte Riese, Johan Ormel, Frank C Verhulst, and Albertine J Oldehinkel. AdolescentsâĂŹ cortisol responses to awakening and social stress; effects of gender, menstrual phase and oral contraceptives. the trails study. *Psychoneuroendocrinology*, 34(6):884–893, 2009.

[9] Allison B Brenner, Marc A Zimmerman, Jose A Bauermeister, and Cleopatra H Caldwell. The physiological expression of living in disadvantaged neighborhoods for youth. *Journal of youth and adolescence*, pages 1–15, 2013.

[10] Stef Buuren and Karin Groothuis-Oudshoorn. Mice: Multivariate imputation by chained equations in r. *Journal of statistical software*, 45(3), 2011.

[11] Edith Chen and Laurel Q Paterson. Neighborhood, family, and subjective socioeconomic status: How do they relate to adolescent health? *Health Psychology*, 25(6):704, 2006.

[12] Yoichi Chida and Andrew Steptoe. Cortisol awakening response and psychosocial factors: a systematic review and meta-analysis. *Biological psychology*, 80(3):265–278, 2009.

[13] Sheldon Cohen, William J Doyle, and Andrew Baum. Socioeconomic status is associated with stress hormones. *Psychosomatic Medicine*, 68(3):414–420, 2006.

[14] Sheldon Cohen, Joseph E Schwartz, Elissa Epel, Clemens Kirschbaum, Steve Sidney, and Teresa Seeman. Socioeconomic status, race, and diurnal cortisol decline in the coronary artery risk development in young adults (cardia) study. *Psychosomatic Medicine*, 68(1):41–50, 2006.

[15] Amy S DeSantis, Emma K Adam, Leah D Doane, Susan Mineka, Richard E Zinbarg, and Michelle G Craske. Racial/ethnic differences in cortisol diurnal rhythms in a community sample of adolescents. *Journal of Adolescent Health*, 41(1):3–13, 2007.

[16] Sally S Dickerson and Margaret E Kemeny. Acute stressors and cortisol responses: a theoretical integration and synthesis of laboratory research. *Psychological bulletin*, 130(3):355, 2004.

[17] Ana V Diez Roux, Catarina I Kiefe, David R Jacobs Jr, Mary Haan, Scott A Jackson, F Javier Nieto, Catherine C Paton, and Richard Schulz. Area characteristics and individual-level socioeconomic position indicators in three population-based epidemiologic studies. *Annals of Epidemiology*, 11(6):395–405, 2001.

[18] Ana V Diez Roux and Christina Mair. Neighborhoods and health. *Annals of the New York Academy of Sciences*, 1186(1):125–145, 2010.

[19] D Phuong Do, Ana V Diez Roux, Anjum Hajat, Amy H Auchincloss, Sharon Stein Merkin, Nalini Ranjit, Steven Shea, and Teresa Seeman. Circadian rhythm of cortisol and neighborhood characteristics in a population-based sample: The multi-ethnic study of atherosclerosis. *Health & place*, 17(2):625–632, 2011.

[20] Julian J Faraway. *Extending the linear model with R: generalized linear, mixed effects and nonparametric regression models*. CRC press, 2005.

[21] Lia CH Fernald and Megan R Gunnar. Poverty-alleviation program participation and salivary cortisol in very low-income children. *Social science & medicine*, 68(12):2180–2189, 2009.

[22] Arline T Geronimus. The weathering hypothesis and the health of african-american women and infants: evidence and speculations. *Ethnicity & disease*, 2(3):207–221, 1991.

[23] Sherita Hill Golden, Gary S Wand, Saurabh Malhotra, Ihab Kamel, and Karen Horton. Reliability of hypothalamic–pituitary–adrenal axis assessment methods for use in population-based studies. *European journal of epidemiology*, 26(7):511–525, 2011.

[24] Sidra Goldman-Mellor, Mark Hamer, and Andrew Steptoe. Early-life stress and recurrent psychological distress over the lifecourse predict divergent cortisol reactivity patterns in adulthood. *Psychoneuroendocrinology*, 37(11):1755–1768, 2012.

[25] WS Gozansky, JS Lynn, ML Laudenslager, and WM Kohrt. Salivary cortisol determined by enzyme immunoassay is preferable to serum total cortisol for

assessment of dynamic hypothalamic–pituitary–adrenal axis activity. *Clinical endocrinology*, 63(3):336–341, 2005.

[26] Per E Gustafsson, Henrik Anckarsäter, Paul Lichtenstein, Nina Nelson, and Per A Gustafsson. Does quantity have a quality all its own? cumulative adversity and up-and down-regulation of circadian salivary cortisol levels in healthy children. *Psychoneuroendocrinology*, 35(9):1410–1415, 2010.

[27] Barbara M Gutteling, Carolina de Weerth, and Jan K Buitelaar. Prenatal stress and children's cortisol reaction to the first day of school. *Psychoneuroendocrinology*, 30(6):541–549, 2005.

[28] Daniel A Hackman, Laura M Betancourt, Nancy L Brodsky, Hallam Hurt, and Martha J Farah. Neighborhood disadvantage and adolescent stress reactivity. *Frontiers in human neuroscience*, 6, 2012.

[29] Anjum Hajat, Ana Diez-Roux, Tracy G Franklin, Teresa Seeman, Sandi Shrager, Nalini Ranjit, Cecilia Castro, Karol Watson, Brisa Sanchez, and Clemens Kirschbaum. Socioeconomic and race/ethnic differences in daily salivary cortisol profiles: The multi-ethnic study of atherosclerosis. *Psychoneuroendocrinology*, 35(6):932–943, 2010.

[30] Carolyn Tucker Halpern, Eric A Whitsel, Brandon Wagner, and Kathleen Mullan Harris. Challenges of measuring diurnal cortisol concentrations in a large population-based field study. *Psychoneuroendocrinology*, 37(4):499–508, 2012.

[31] Mark Hamer, Katie O'Donnell, Avijit Lahiri, and Andrew Steptoe. Salivary cortisol responses to mental stress are associated with coronary artery calcification in healthy men and women. *European heart journal*, 31(4):424–429, 2010.

[32] Claire Henderson, Ana V Diez Roux, David R Jacobs, Catarina I Kiefe, Delia West, and David R Williams. Neighbourhood characteristics, individual level socioeconomic factors, and depressive symptoms in young adults: the cardia study. *Journal of Epidemiology and Community Health*, 59(4):322–328, 2005.

[33] Miguel A Hernán, Sonia Hernandez-Diaz, and James M Robins. A structural approach to selection bias. *Epidemiology*, 15(5):615–625, 2004.

[34] Kim Hinkelmann, Steffen Moritz, Johannes Botzenhardt, Kirsten Riedesel, Klaus Wiedemann, Michael Kellner, and Christian Otte. Cognitive impairment in major depression: association with salivary cortisol. *Biological psychiatry*, 66(9):879–885, 2009.

[35] Daniel E Ho, Kosuke Imai, Gary King, and Elizabeth A Stuart. Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political analysis*, 15(3):199–236, 2007.

[36] Rebecca A Karb, Michael R Elliott, Jennifer B Dowd, and Jeffrey D Morenoff. Neighborhood-level stressors, social support, and diurnal patterns of cortisol: The chicago community adult health study. *Social science & medicine*, 75(6):1038–1047, 2012.

[37] Shona J Kelly, Robert Young, Helen Sweeting, Joachim E Fischer, and Patrick West. Levels and confounders of morning cortisol collected from adolescents in a naturalistic (school) setting. *Psychoneuroendocrinology*, 33(9):1257–1268, 2008.

[38] Ronald C Kessler, Shelli Avenevoli, E Jane Costello, Jennifer Greif Green, Michael J Gruber, Steven Heeringa, Kathleen R Merikangas, Beth-Ellen Pennell, Nancy A Sampson, and Alan M Zaslavsky. Design and field procedures

in the us national comorbidity survey replication adolescent supplement (ncs-a). *International Journal of Methods in Psychiatric Research*, 18(2):69–83, 2009.

[39] Ronald C Kessler, Shelli Avenevoli, E Jane Costello, Jennifer Greif Green, Michael J Gruber, Steven Heeringa, Kathleen R Merikangas, Beth-Ellen Pennell, Nancy A Sampson, and Alan M Zaslavsky. National comorbidity survey replication adolescent supplement (NCS-A): II. overview and design. *Journal of the American Academy of Child & Adolescent Psychiatry*, 48(4):380–385, 2009.

[40] Ronald C Kessler, Shelli Avenevoli, Jennifer Green, Michael J Gruber, Margaret Guyer, Yulei He, Robert Jin, Joan Kaufman, Nancy A Sampson, Alan M Zaslavsky, et al. National comorbidity survey replication adolescent supplement (ncs-a): Iii. concordance of< i> dsm-iv</i>/cidi diagnoses with clinical reassessments. *Journal of the American Academy of Child & Adolescent Psychiatry*, 48(4):386–399, 2009.

[41] Clemens Kirschbaum and Dirk H Hellhammer. Salivary cortisol in psychoneuroendocrine research: recent developments and applications. *Psychoneuroendocrinology*, 19(4):313–333, 1994.

[42] Clemens Kirschbaum, K-M Pirke, and Dirk H Hellhammer. The âĂŸtrier social stress testâĂŹ–a tool for investigating psychobiological stress responses in a laboratory setting. *Neuropsychobiology*, 28(1-2):76–81, 1993.

[43] Tama Leventhal and Jeanne Brooks-Gunn. The neighborhoods they live in: the effects of neighborhood residence on child and adolescent outcomes. *Psychological bulletin*, 126(2):309, 2000.

[44] D Liu, C Caldji, S Sharma, PM Plotsky, and MJ Meaney. Influence of neonatal rearing conditions on stress-induced adrenocorticotropin responses and norepinepherine release in the hypothalamic paraventricular nucleus. *Journal of neuroendocrinology*, 12(1):5–12, 2000.

[45] Dong Liu, Josie Diorio, Beth Tannenbaum, Christian Caldji, Darlene Francis, Alison Freedman, Shakti Sharma, Deborah Pearson, Paul M Plotsky, and Michael J Meaney. Maternal care, hippocampal glucocorticoid receptors, and hypothalamic-pituitary-adrenal responses to stress. *Science*, 277(5332):1659–1662, 1997.

[46] William R Lovallo, Noha H Farag, Kristen H Sorocco, Andrew J Cohoon, and Andrea S Vincent. Lifetime adversity leads to blunted stress axis reactivity: Studies from the oklahoma family health patterns project. *Biological psychiatry*, 71(4):344–349, 2012.

[47] Thomas Lumley. Analysis of complex survey samples. *Journal of Statistical Software*, 9(1):1–19, 2004.

[48] Harriet L MacMillan, Katholiki Georgiades, Eric K Duku, Alison Shea, Meir Steiner, Anne Niec, Masako Tanaka, Susan Gensey, Sandra Spree, Emily Vella, et al. Cortisol response to stress in female youths exposed to childhood maltreatment: results of the youth mood project. *Biological psychiatry*, 66(1):62–68, 2009.

[49] Deborah Mangold, Gary Wand, Martin Javors, and James Mintz. Acculturation, childhood trauma and the cortisol awakening response in mexican–american adults. *Hormones and behavior*, 58(4):637–646, 2010.

[50] Christina Gamache Martin, Jacqueline Bruce, and Philip A Fisher. Racial and ethnic differences in diurnal cortisol rhythms in preadolescents: The role of parental psychosocial risk and monitoring. *Hormones and behavior*, 61(5):661–668, 2012.

[51] Bruce S McEwen. Physiology and neurobiology of stress and adaptation: central role of the brain. *Physiological reviews*, 87(3):873–904, 2007.

[52] Michael J Meaney, Beth Tannenbaum, Darlene Francis, Seema Bhatnagar, Nola Shanks, Victor Viau, Dajan O'Donnell, and Paul M Plotsky. Early environmental programming hypothalamic-pituitary-adrenal responses to stress. In *Seminars in Neuroscience*, volume 6, pages 247–259. Elsevier, 1994.

[53] Kathleen R Merikangas, Shelli Avenevoli, E Jane Costello, Doreen Koretz, and Ronald C Kessler. National comorbidity survey replication adolescent supplement (NCS-A): I. background and measures. *Journal of the American Academy of Child & Adolescent Psychiatry*, 48(4):367–379, 2009.

[54] Aydin Nazmi, Ana Diez Roux, Nalini Ranjit, Teresa E Seeman, and Nancy S Jenny. Cross-sectional and longitudinal associations of neighborhood characteristics with inflammatory markers: findings from the multi-ethnic study of atherosclerosis. *Health & place*, 16(6):1104–1112, 2010.

[55] Cheryl K Nordstrom, Ana V Diez Roux, Richard Schulz, Mary N Haan, Sharon A Jackson, and Jennifer L Balfour. Socioeconomic position and incident mobility impairment in the cardiovascular health study. *BMC geriatrics*, 7(1):11, 2007.

[56] Anu-Katriina Pesonen, Katri Räikkönen, Kimmo Feldt, Kati Heinonen, Clive Osmond, David IW Phillips, David JP Barker, Johan G Eriksson, and Eero Ka-

jantie. Childhood separation experience predicts hpa axis hormonal responses in late adulthood: a natural experiment of world war ii. *Psychoneuroendocrinology*, 35(5):758–767, 2010.

[57] Christoph Randler and Steffen Schaal. Morningness–eveningness, habitual sleep-wake variables and cortisol level. *Biological psychology*, 85(1):14–18, 2010.

[58] Rena L Repetti, Shelley E Taylor, and Teresa E Seeman. Risky families: family social environments and the mental and physical health of offspring. *Psychological bulletin*, 128(2):330, 2002.

[59] Paul R Rosenbaum and Donald B Rubin. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55, 1983.

[60] JGM Rosmalen, AJ Oldehinkel, J Ormel, AF De Winter, JK Buitelaar, and FC Verhulst. Determinants of salivary cortisol levels in 10–12 year old children; a population-based study of individual differences. *Psychoneuroendocrinology*, 30(5):483–495, 2005.

[61] Ana V Diez Roux, Luisa N Borrell, Mary Haan, Sharon A Jackson, and Richard Schultz. Neighbourhood environments and mortality in an elderly cohort: results from the cardiovascular health study. *Journal of epidemiology and community health*, 58(11):917–923, 2004.

[62] Donald B Rubin. *Multiple imputation for nonresponse in surveys*, volume 307. Wiley. com, 2009.

[63] Donald B Rubin and Neal Thomas. Combining propensity score matching with

additional adjustments for prognostic covariates. *Journal of the American Statistical Association*, 95(450):573–585, 2000.

[64] Kara E Rudolph, Gary S Wand, Elizabeth A Stuart, Thomas A Glass, Andrea H Marques, Roman Duncko, and Kathleen R Merikangas. The association between cortisol and neighborhood disadvantage in a us population-based sample of adolescents. *Health & Place*, 25:68âĂŞ77, 2014.

[65] Robert M Sapolsky, Gordon Gurley, David Demarest, and Entreprises internationales Télé-Film. *Why zebras don't get ulcers*. Times Books New York, 2004.

[66] Teresa E Seeman, Eileen Crimmins, Mei-Hua Huang, Burton Singer, Alexander Bucur, Tara Gruenewald, Lisa F Berkman, and David B Reuben. Cumulative biological risk and socio-economic differences in mortality: Macarthur studies of successful aging. *Social science & medicine*, 58(10):1985–1997, 2004.

[67] Andrew Steptoe, Sabine R Kunz-Ebrecht, Caroline Wright, and Pamela J Feldman. Socioeconomic position and cardiovascular and neuroendocrine responses following cognitive challenge in old age. *Biological psychology*, 69(2):149–166, 2005.

[68] Elizabeth A Stuart, Melissa Azur, Constantine Frangakis, and Philip Leaf. Multiple imputation with large data sets: a case study of the children's mental health initiative. *American journal of epidemiology*, 169(9):1133–1139, 2009.

[69] Tyler J VanderWeele and Onyebuchi A Arah. Bias formulas for sensitivity analysis of unmeasured confounding for general outcomes, treatments, and confounders. *Epidemiology*, 22(1):42–52, 2011.

[70] RF Vining, RA McGinley, JJ Maksvytis, and KIAN Y Ho. Salivary cortisol: a better measure of adrenal cortical function than serum cortisol. *Annals of Clinical Biochemistry*, 20:329–335, 1983.

[71] Michelle M Wirth, Kathryn M Welsh, and Oliver C Schultheiss. Salivary cortisol changes in humans after winning or losing a dominance contest depend on implicit power motivation. *Hormones and Behavior*, 49(3):346–352, 2006.

[72] DB Woodside, K Winter, and S Fisman. Salivary cortisol in children: correlations with serum values and effect of psychotropic drug administration. *Canadian journal of psychiatry. Revue canadienne de psychiatrie*, 36(10):746, 1991.

References for Chapter 4

[1] Heejung Bang and James M Robins. Doubly robust estimation in missing data and causal inference models. *Biometrics*, 61(4):962–973, 2005.

[2] Weihua Cao, Anastasios A Tsiatis, and Marie Davidian. Improving efficiency and robustness of the doubly robust estimator for a population mean with incomplete data. *Biometrika*, 96(3):723–734, 2009.

[3] Paul Chaffee, Alan E Hubbard, and Mark L van der Laan. Permutation-based pathway testing using the super learner algorithm. *U.C. Berkeley Division of Biostatistics Working Paper Series*, 263, 2010.

[4] Stephen R Cole and Elizabeth A Stuart. Generalizing evidence from randomized clinical trials to target populations the ACTG 320 trial. *American Journal of Epidemiology*, 172(1):107–115, 2010.

[5] Rajeev H Dehejia and Sadek Wahba. Propensity score-matching methods for nonexperimental causal studies. *Review of Economics and Statistics*, 84(1):151–161, 2002.

[6] Ana V Diez Roux, Catarina I Kiefe, David R Jacobs Jr, Mary Haan, Scott A Jackson, F Javier Nieto, Catherine C Paton, and Richard Schulz. Area character-

istics and individual-level socioeconomic position indicators in three population-based epidemiologic studies. *Annals of Epidemiology*, 11(6):395–405, 2001.

[7] Susan Gruber and Mark van der Laan. tmle: An R package for targeted maximum likelihood estimation. *Journal of Statistical Software*, 51.

[8] Daniel G Horvitz and Donovan J Thompson. A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47(260):663–685, 1952.

[9] Joseph DY Kang and Joseph L Schafer. Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data. *Statistical Science*, 22(4):523–539, 2007.

[10] Ronald C Kessler, Shelli Avenevoli, E Jane Costello, Jennifer Greif Green, Michael J Gruber, Steven Heeringa, Kathleen R Merikangas, Beth-Ellen Pennell, Nancy A Sampson, and Alan M Zaslavsky. Design and field procedures in the us national comorbidity survey replication adolescent supplement (ncs-a). *International Journal of Methods in Psychiatric Research*, 18(2):69–83, 2009.

[11] Ronald C Kessler, Shelli Avenevoli, E Jane Costello, Jennifer Greif Green, Michael J Gruber, Steven Heeringa, Kathleen R Merikangas, Beth-Ellen Pennell, Nancy A Sampson, and Alan M Zaslavsky. National comorbidity survey replication adolescent supplement (NCS-A): II. overview and design. *Journal of the American Academy of Child & Adolescent Psychiatry*, 48(4):380–385, 2009.

[12] Ronald C Kessler, Shelli Avenevoli, Jennifer Green, Michael J Gruber, Margaret Guyer, Yulei He, Robert Jin, Joan Kaufman, Nancy A Sampson, Alan M Zaslavsky, et al. National comorbidity survey replication adolescent supplement

(NCS-A): III. concordance of DSM-IV CIDI diagnoses with clinical reassessments. *Journal of the American Academy of Child & Adolescent Psychiatry*, 48(4):386–399, 2009.

[13] Jae Kwang Kim and David Haziza. Doubly robust inference with missing data in survey sampling. In *Joint Statistical Meetings*. Section on Survey Research Methods, 2010.

[14] Selene Leon, Anastasios A Tsiatis, and Marie Davidian. Semiparametric estimation of treatment effect in a pretest-posttest study. *Biometrics*, 59(4):1046–1055, 2003.

[15] Bruce S McEwen. Physiology and neurobiology of stress and adaptation: central role of the brain. *Physiological reviews*, 87(3):873–904, 2007.

[16] Kathleen R Merikangas, Shelli Avenevoli, E Jane Costello, Doreen Koretz, and Ronald C Kessler. National comorbidity survey replication adolescent supplement (NCS-A): I. background and measures. *Journal of the American Academy of Child & Adolescent Psychiatry*, 48(4):367–379, 2009.

[17] Lynne C Messer, J Michael Oakes, and Susan Mason. Effects of socioeconomic and racial residential segregation on preterm birth: a cautionary tale of structural confounding. *American Journal of Epidemiology*, 171(6):664–673, 2010.

[18] Kelly L Moore, Romain Neugebauer, Mark J van der Laan, and Ira B Tager. Causal inference in epidemiological studies with strong confounding. *Statistics in Medicine*, 31(13):1380–1404, 2012.

[19] Judea Pearl and Elias Bareinboim. Transportability of causal and statistical

relations: A formal approach. In *Data Mining Workshops (ICDMW), 2011 IEEE 11th International Conference on*, pages 540–547. IEEE, 2011.

[20] Maya L Petersen, Kristin E Porter, Susan Gruber, Yue Wang, and Mark J van der Laan. Diagnosing and responding to violations in the positivity assumption. *Statistical Methods in Medical Research*, 2010.

[21] Maya L Petersen, Joshua Schwab, Susan Gruber, Nello Blaser, Michael Schomaker, and Mark J van der Laan. Targeted maximum likelihood estimation for dynamic and static longitudinal marginal structural working models. *U.C. Berkeley Division of Biostatistics Working Paper Series*, 312, 2013.

[22] Kristin E Porter, Susan Gruber, Mark J van der Laan, and Jasjeet S Sekhon. The relative performance of targeted maximum likelihood estimators. *The International Journal of Biostatistics*, 7(1):1–34, 2011.

[23] James Robins, Mariela Sued, Quanhong Lei-Gomez, and Andrea Rotnitzky. Comment: Performance of double-robust estimators when "inverse probability" weights are highly variable. *Statistical Science*, 22(4):544–559, 2007.

[24] James M Robins, Andrea Rotnitzky, and Lue Ping Zhao. Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association*, 90(429):106–121, 1995.

[25] Paul R Rosenbaum and Donald B Rubin. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55, 1983.

[26] Kara E Rudolph, Iván Díaz, Michael Rosenblum, and Elizabeth A Stuart. Estimating population treatment effects from a survey sub-sample. 2013.

[27] Kara E Rudolph, Gary S Wand, Elizabeth A Stuart, Thomas A Glass, Andrea H Marques, Roman Duncko, and Kathleen R Merikangas. The association between cortisol and neighborhood disadvantage in a us population-based sample of adolescents. *Health & Place*, 25:68âĂŞ77, 2014.

[28] Daniel O Scharfstein, Andrea Rotnitzky, and James M Robins. Adjusting for nonignorable drop-out using semiparametric nonresponse models. *Journal of the American Statistical Association*, 94(448):1096–1120, 1999.

[29] Jonathan M Snowden, Sherri Rose, and Kathleen M Mortimer. Implementation of g-computation on a simulated data set: Demonstration of a causal inference technique. *American Journal of Epidemiology*, 173(7):731, 2011.

[30] Elizabeth A Stuart, Stephen R Cole, Catherine P Bradshaw, and Philip J Leaf. The use of propensity scores to assess the generalizability of results from randomized trials. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 174(2):369–386, 2011.

[31] Mark J van der Laan. Targeted maximum likelihood based causal inference: Part i. *The International Journal of Biostatistics*, 6(2), 2010.

[32] Mark J van der Laan. Targeted maximum likelihood based causal inference: Part ii. *The International Journal of Biostatistics*, 6(2), 2010.

[33] Mark J van der Laan and Daniel Rubin. Targeted maximum likelihood learning. *The International Journal of Biostatistics*, 2(1), 2006.

[34] Yue Wang, Maya L Petersen, David Bangsberg, and Mark J van der Laan. Diagnosing bias in the inverse probability of treatment weighted estimator resulting

from violation of experimental treatment assignment. *U.C. Berkeley Division of Biostatistics Working Paper Series*, 211, 2006.

References for Chapter 5

[1] Karen Bandeen-Roche, Diana L Miglioretti, Scott L Zeger, and Paul J Rathouz. Latent variable regression for multiple discrete outcomes. *Journal of the American Statistical Association*, 92(440):1375–1386, 1997.

[2] Kenneth A Bollen and Gerhard Arminger. Observational residuals in factor analysis and structural equation models. *Sociological methodology*, 21:235–262, 1991.

[3] Yoichi Chida and Andrew Steptoe. Cortisol awakening response and psychosocial factors: a systematic review and meta-analysis. *Biological psychology*, 80(3):265–278, 2009.

[4] Stephen R Cole and Miguel A Hernán. Constructing inverse probability weights for marginal structural models. *American journal of epidemiology*, 168(6):656–664, 2008.

[5] Eva H DuGoff, Megan Schuler, and Elizabeth A Stuart. Generalizing observational study results: Applying propensity score methods to complex surveys. *Health services research*, 2013.

[6] Alan Feingold, Stacey S Tiberio, and Deborah M Capaldi. New approaches for

examining associations with latent categorical variables: Applications to substance abuse and aggression. 2013.

[7] Sandro Galea. The urban brain: new directions in research exploring the relation between cities and mood–anxiety disorders. *Depression and anxiety*, 28(10):857–862, 2011.

[8] Nancy A Gonzales, Stefany Coxe, Mark W Roosa, Rebecca MB White, George P Knight, Katharine H Zeiders, and Delia Saenz. Economic hardship, neighborhood context, and parenting: Prospective effects on mexican–american adolescentâĂŹs mental health. *American journal of community psychology*, 47(1-2):98–113, 2011.

[9] WS Gozansky, JS Lynn, ML Laudenslager, and WM Kohrt. Salivary cortisol determined by enzyme immunoassay is preferable to serum total cortisol for assessment of dynamic hypothalamic–pituitary–adrenal axis activity. *Clinical endocrinology*, 63(3):336–341, 2005.

[10] Sander Greenland and Hal Morgenstern. Ecological bias, confounding, and effect modification. *International journal of epidemiology*, 18(1):269–274, 1989.

[11] Anjum Hajat, Ana Diez-Roux, Tracy G Franklin, Teresa Seeman, Sandi Shrager, Nalini Ranjit, Cecilia Castro, Karol Watson, Brisa Sanchez, and Clemens Kirschbaum. Socioeconomic and race/ethnic differences in daily salivary cortisol profiles: The multi-ethnic study of atherosclerosis. *Psychoneuroendocrinology*, 35(6):932–943, 2010.

[12] Carolyn Tucker Halpern, Eric A Whitsel, Brandon Wagner, and Kathleen Mullan Harris. Challenges of measuring diurnal cortisol concentrations in a large population-based field study. *Psychoneuroendocrinology*, 37(4):499–508, 2012.

[13] Mark Hamer, Katie O'Donnell, Avijit Lahiri, and Andrew Steptoe. Salivary cortisol responses to mental stress are associated with coronary artery calcification in healthy men and women. *European heart journal*, 31(4):424–429, 2010.

[14] Miguel Hernan and James Robins. *Causal inference*. Chapman Hall, CRC, 2013.

[15] Kosuke Imai and Marc Ratkovic. Estimating treatment effect heterogeneity in randomized program evaluation. *The Annals of Applied Statistics*, 7(1):443–470, 2013.

[16] Ronald C Kessler, Shelli Avenevoli, Jennifer Green, Michael J Gruber, Margaret Guyer, Yulei He, Robert Jin, Joan Kaufman, Nancy A Sampson, Alan M Zaslavsky, et al. National comorbidity survey replication adolescent supplement (ncs-a): Iii. concordance of< i> dsm-iv</i>/cidi diagnoses with clinical reassessments. *Journal of the American Academy of Child & Adolescent Psychiatry*, 48(4):386–399, 2009.

[17] Clemens Kirschbaum and Dirk H Hellhammer. Salivary cortisol in psychoneuroendocrine research: recent developments and applications. *Psychoneuroendocrinology*, 19(4):313–333, 1994.

[18] Jeffrey R Kling, Jeffrey B Liebman, and Lawrence F Katz. Experimental analysis of neighborhood effects. *Econometrica*, 75(1):83–119, 2007.

[19] Nancy Krieger, Jarvis T Chen, Pamela D Waterman, Mah-Jabeen Soobader, SV Subramanian, and Rosa Carson. Geocoding and monitoring of us socioeconomic inequalities in mortality and cancer incidence: does the choice of area-based measure and geographic level matter? the public health disparities geocoding project. *American journal of epidemiology*, 156(5):471–482, 2002.

[20] Tama Leventhal and Jeanne Brooks-Gunn. Moving to opportunity: an experimental study of neighborhood effects on mental health. *American Journal of Public Health*, 93(9):1576–1582, 2003.

[21] Daniel F McCaffrey, JR Lockwood, and CLAUDE M SETODJI. Inverse probability weighting with error-prone covariates. 2011.

[22] Bruce S McEwen and Teresa Seeman. Protective and damaging effects of mediators of stress: elaborating and testing the concepts of allostasis and allostatic load. *Annals of the New York Academy of Sciences*, 896(1):30–47, 1999.

[23] Daniel L Millimet. *The elephant in the corner: a cautionary tale about measurement error in treatment effects models*, volume 27. Emerald Group Publishing Limited, 2011.

[24] Tenko Raykov. Propensity score analysis with fallible covariates a note on a latent variable modeling approach. *Educational and Psychological Measurement*, 72(5):715–733, 2012.

[25] James Robins, Mariela Sued, Quanhong Lei-Gomez, and Andrea Rotnitzky. Comment: Performance of double-robust estimators when "inverse probability" weights are highly variable. *Statistical Science*, 22(4):544–559, 2007.

[26] Paul R Rosenbaum. *Observational studies*. Springer, 2 edition, 2002.

[27] Ana V Diez Roux, Luisa N Borrell, Mary Haan, Sharon A Jackson, and Richard Schultz. Neighbourhood environments and mortality in an elderly cohort: results from the cardiovascular health study. *Journal of epidemiology and community health*, 58(11):917–923, 2004.

[28] Donald B Rubin. Comment: Which ifs have causal answers. *Journal of the American Statistical Association*, 81(396):961–962, 1986.

[29] Teresa E Seeman, Eileen Crimmins, Mei-Hua Huang, Burton Singer, Alexander Bucur, Tara Gruenewald, Lisa F Berkman, and David B Reuben. Cumulative biological risk and socio-economic differences in mortality: Macarthur studies of successful aging. *Social science & medicine*, 58(10):1985–1997, 2004.

[30] Ronald L Simons, Velma Murry, Vonnie McLoyd, Kuei-Hsiu Lin, Carolyn Cutrona, Rand D Conger, et al. Discrimination, crime, ethnic identity, and parenting as correlates of depressive symptoms among african american children: A multilevel analysis. *Development and psychopathology*, 14(2):371–393, 2002.

[31] Natalie Slopen, Garrett M Fitzmaurice, David R Williams, and Stephen E Gilman. Common patterns of violence experiences and depression and anxiety among adolescents. *Social psychiatry and psychiatric epidemiology*, 47(10):1591–1605, 2012.

[32] Michael E Sobel. What do randomized studies of housing mobility demonstrate? *Journal of the American Statistical Association*, 101(476), 2006.

[33] Peter M Steiner, Thomas D Cook, and William R Shadish. On the importance of reliable covariate measurement in selection bias adjustments using propensity scores. *Journal of Educational and Behavioral Statistics*, 36(2):213–236, 2011.

[34] Elizabeth A Stuart, Melissa Azur, Constantine Frangakis, and Philip Leaf. Multiple imputation with large data sets: a case study of the children's mental health initiative. *American journal of epidemiology*, 169(9):1133–1139, 2009.

[35] Til Stürmer, Sebastian Schneeweiss, Jerry Avorn, and Robert J Glynn. Adjusting effect estimates for unmeasured confounding with validation data using propensity score calibration. *American journal of epidemiology*, 162(3):279–289, 2005.

[36] Eric J Tchetgen Tchetgen and Tyler J VanderWeele. On causal inference in the presence of interference. *Statistical Methods in Medical Research*, 21(1):55–75, 2012.

[37] Stef Van Buuren. Multiple imputation of discrete and continuous data by fully conditional specification. *Statistical methods in medical research*, 16(3):219–242, 2007.

[38] Mark J van der Laan. Causal inference for networks. 2012.

[39] Tyler J VanderWeele and Onyebuchi A Arah. Bias formulas for sensitivity analysis of unmeasured confounding for general outcomes, treatments, and confounders. *Epidemiology*, 22(1):42–52, 2011.

[40] RF Vining, RA McGinley, JJ Maksvytis, and KIAN Y Ho. Salivary cortisol: a better measure of adrenal cortical function than serum cortisol. *Annals of Clinical Biochemistry*, 20:329–335, 1983.

[41] William J Wilson. *When work disappears: The world of the new urban poor*. Random House Digital, Inc., 1997.

[42] Michelle M Wirth, Kathryn M Welsh, and Oliver C Schultheiss. Salivary cortisol changes in humans after winning or losing a dominance contest depend on implicit power motivation. *Hormones and Behavior*, 49(3):346–352, 2006.

[43] Hans-Ulrich Wittchen. Reliability and validity studies of the who-composite international diagnostic interview (cidi): a critical review. *Journal of psychiatric research*, 28(1):57–84, 1994.

[44] DB Woodside, K Winter, and S Fisman. Salivary cortisol in children: correlations with serum values and effect of psychotropic drug administration. *Canadian journal of psychiatry. Revue canadienne de psychiatrie*, 36(10):746, 1991.

[45] Yange Xue, Tama Leventhal, Jeanne Brooks-Gunn, and Felton J Earls. Neighborhood residence and mental health problems of 5-to 11-year-olds. *Archives of General Psychiatry*, 62(5):554, 2005.

[46] Elaine L Zanutto. A comparison of propensity score and linear regression analysis of complex survey data. *Journal of Data Science*, 4(1):67–91, 2006.

References for Appendix A

[1] J Douglas Bremner. Neuroimaging studies in post-traumatic stress disorder. *Current Psychiatry Reports*, 4:254–263, 2002.

[2] Sheldon Cohen. Psychosocial models of the role of social support in the etiology of physical disease. *Health psychology*, 7(3):269, 1988.

[3] Holly Foster and Jeanne Brooks-Gunn. Toward a stress process model of childrenâĂŹs exposure to physical family and community violence. *Clinical child and family psychology review*, 12(2):71–94, 2009.

[4] Phillip L Hammack, Maryse H Richards, Zupei Luo, Emily S Edlynn, and Kevin Roy. Social support factors as moderators of community violence exposure among inner-city african american young adolescents. *Journal of Clinical Child and Adolescent Psychology*, 33(3):450–462, 2004.

[5] Keisuke Hirano and Guido W Imbens. The propensity score with continuous treatments. *Applied Bayesian modeling and causal inference from incomplete-data perspectives*, 226164:73–84, 2004.

[6] Christina Mair, Ana V Diez Roux, and Jeffrey D Morenoff. Neighborhood stres-

sors and social support as predictors of depressive symptoms in the chicago community adult health study. *Health & place*, 16(5):811–819, 2010.

[7] Bruce S McEwen. Physiology and neurobiology of stress and adaptation: central role of the brain. *Physiological reviews*, 87(3):873–904, 2007.

[8] Roger K Pitman. Hippocampal diminution in ptsd: more (or less?) than meets the eye. *Hippocampus*, 11(2):73–74, 2001.

[9] Teresa E Seeman, Linda Fagan Dubin, and Melvin Seeman. Religiosity/spirituality and health: A critical review of the evidence for biological pathways. *American Psychologist*, 58(1):53, 2003.

[10] Yvette I Sheline, Milan Sanghavi, Mark A Mintun, and Mokhtar H Gado. Depression duration but not age predicts hippocampal volume loss in medically healthy women with recurrent major depression. *The Journal of Neuroscience*, 19(12):5034–5043, 1999.

[11] Monica N Starkman, Stephen S Gebarski, Stanley Berent, and David E Schteingart. Hippocampal formation volume, memory dysfunction, and cortisol levels in patients with cushing's syndrome. *Biological psychiatry*, 32(9):756–765, 1992.

[12] Sophie A Vreeburg, Witte JG Hoogendijk, Johannes van Pelt, Roel H DeRijk, Jolanda Verhagen, Richard van Dyck, Johannes H Smit, Frans G Zitman, and Brenda WJH Penninx. Major depressive disorder and hypothalamic-pituitary-adrenal axis activity: results from a large cohort study. *Archives of General Psychiatry*, 66(6):617, 2009.

[13] Elizabeth A Young, Roger F Haskett, Leon Grunhaus, Atul Pande, Virginia Murphy Weinberg, Stanley J Watson, and Huda Akil. Increased evening activation of the hypothalamic-pituitary-adrenal axis in depressed patients. *Archives of general psychiatry*, 51(9):701, 1994.

Curriculum Vitae

Kara E. Rudolph
Born January 30, 1983 in Bay City, Michigan, USA
Current Address:
1106 E. Capitol St. NE Apt 2
Washington, DC 20002
krudolph@jhsph.edu

Education

2010-2014	**Johns Hopkins School of Public Health, Baltimore, MD** Ph.D. Epidemiology, - *expected*
2013-2013	**Johns Hopkins School of Public Health, Baltimore, MD** M.H.S. Biostatistics Date of graduation: December 2013
2007-2008	**Johns Hopkins School of Public Health, Baltimore, MD** M.P.H. Date of graduation: May 2008
2006	**University of Pennsylvania, Philadelphia, PA** Graduate course in biostatistics.
2001-2005	**University of Michigan, Ann Arbor, MI** B.S. Highest honors in biochemistry. Date of graduation: May 2005

Professional Experience

Research

2013-	**Georgetown University** **Washington, DC** Center for Child and Human Development Statistical consultant for Partners for a Healthy Baby Evaluation
2012-2013	**Johns Hopkins School of Public Health** **Baltimore, MD** Department of Mental Health Graduate research assistant to Dr. Elizabeth Stuart Developed conference material on statistical methods for missing data, developed material for Causal Inference course.
2010-2013	**Johns Hopkins School of Public Health** **Baltimore, MD** Department of Epidemiology Graduate research assistant to Drs. Derek Cummings and Justin Lessler Performed systematic reviews of the incubation periods of multiple viruses.

Teaching Experience

Johns Hopkins Bloomberg School of Public Health
Baltimore, MD
Teaching assistant
Department of Biostatistics

2013	Spatial Analysis and GIS I
2013	Spatial Analysis and GIS II

Department of Epidemiology

2012	Methodological challenges in epidemiological research
2011	Doctoral Seminar in epidemiology
2011-2013	Fundamentals of epidemiology and biostatistics

Publications

Rudolph KE, Wand GS, Stuart EA, Glass TA, Marques AH, Duncko R, Merikangas KR. The association between cortisol characteristics and neighborhood disadvantage in a population-based sample of U.S. adolescents. *Health Place*. 2014; 25: 68–77.

Rudolph KE, Stuart EA, Glass TA, Merikangas KR. Neighborhood disadvantage in

context: the influence of urbanicity on the association between neighborhood disadvantage and adolescent emotional disorders. *Soc Psychiatry Psychiatr Epidemiol*. In press.

Rudolph KE, Glass TA, Crum RM, Schwartz BS. Neighborhood psychosocial hazards and binge drinking among late middle-aged adults. *J Urban Health*. 2013; 90(5): 970-82

Rudolph KE, Lessler J, Moloney R, Kmush B, Cummings DAT. Incubation periods of mosquito-borne infections: a systematic review. *Am J Trop Med Hyg*. In press.

Lee RM, Lessler J, Lee RA, Rudolph KE, Reich NG, Perl TM, Cummings DAT. Incubation periods of viral gastroenteritis: a systematic review. *BMC Infect Dis*. 2013; 13: 446.

Azman AS, Rudolph KE, Cummings DAT, Lessler J. The incubation period of cholera: a systematic review. May 2008. *J Infect*. 2013; 66(5): 432-8.

Presentations and Abstracts

Abstracts *(first author only)*

Rudolph KE, Wand GS, Stuart EA, Glass TA, Marques AH, Duncko R, Merikangas KR. The association between cortisol characteristics and neighborhood disadvantage in a population-based sample of U.S. adolescents. Poster at the Society for Epidemiologic Research Conference, Boston, MA, 2013.

Rudolph KE, Stuart EA, Glass TA, Merikangas, KR. Using propensity score subclassification to examine the association between neighborhood disadvantage and adolescent mental health. Poster at the Society for Epidemiologic Research Conference, Minneapolis, MN, 2012.

Rudolph KE, Glass TA, Schwartz BS. Neighborhood psychosocial hazards and binge drinking among older adults. Poster at the Society for Epidemiologic Research Conference, Minneapolis, MN, 2012.

Rudolph K, Gabor V, Morgan R. Environmental supports and barriers to physical activity in D.C. child care centers. Oral presentation at the Annual American Public Health Association Conference, Denver, CO, 2010.

Professional Activities

Reviewer, *Journal of Epidemiology and Community Health* 2013

Awards

2014	Student Paper Competition Winner, American Statistical Association, Social Statistics/Government/ Survey Research Methods Sections
2012-2014	Sommer Scholar, Johns Hopkins School of Public Health
2010-2012	Intramural Research Training Award Fellow, National Institutes of Mental Health
2008	Delta Omega Honor Society, Johns Hopkins School of Public Health
2003-2005	James B. Angell Scholar, University of Michigan
2001-2005	University Honors, University of Michigan